DOES NOT CIRCULATE

Vol. 11 No. 2

UNIV. OF MICHIGAN
DEC 16 1957
PUBLIC HEALTH LIBRARY

APRIL, 1957

BRITISH JOURNAL OF PREVENTIVE & SOCIAL MEDICINE

EDITED BY
THOMAS McKEOWN AND IAN TAYLOR

WITH THE ASSISTANCE OF

W. MELVILLE ARNOTT
F. A. E. CREW
A. BRADFORD HILL
LANCELOT HOGBEN
W. J. E. JESSOP
AUBREY LEWIS
W. P. D. LOGAN
J. N. MORRIS
J. A. FRASER ROBERTS
A. C. STEVENSON
ALICE M. STEWART
THE EDITOR OF THE *British Medical Journal*

CONTENTS

PAGE

The Stature and Weight of Men from England and Wales in 1941. E. M. B. Clements and Kathleen G. Pickett 51

Random Selection in the Therapeutic Trial. R. F. Wrighton 61

A Note on the Practical Interpretation of 2 × 2 Tables. J. H. Edwards 73

A Critical Examination of the Reputed Primary Influence of ABO Phenotype on Fertility and Sex Ratio. J. H. Edwards 79

Recent Trends of Mortality from Violence. W. J. Martin 90

An Examination of Fertility of Women following Pregnancy according to Height and Weight. Thomas McKeown and R. G. Record 102

LONDON
BRITISH MEDICAL ASSOCIATION
TAVISTOCK SQUARE, W.C.1

YEARLY SUBSCRIPTION (4 NUMBERS) £2 2s. 0d. U.S.A. $7.00 SINGLE NUMBER 12s. 6d.

NOTICE TO CONTRIBUTORS

THE BRITISH JOURNAL OF PREVENTIVE AND SOCIAL MEDICINE is intended for the publication of original work in the field. Contributions in the English language from workers of any nationality may be considered for publication.

As understood by those who are responsible for this Journal, social medicine is that branch of science which is concerned with: (*a*) biological needs, interactions, disabilities, and potentialities of human beings living in social aggregates; (*b*) numerical, structural, and functional changes of human populations in their biological and medical aspects. To a large extent its methods must necessarily be statistical, involving the use of numerical data obtained either from official sources or from special field investigations, and interpreted in the light of established findings of the laboratory and of the clinic. Social medicine takes within its province the study of all environmental agencies, living and non-living, relevant to health and efficiency, also fertility and population genetics, norms and ranges of variation with respect to individual differences, and investigations directed to the assessment of a regimen of positive health.

All papers submitted for publication and all other editorial matters should be addressed to The Editors, BRITISH JOURNAL OF PREVENTIVE AND SOCIAL MEDICINE, c/o British Medical Association, Tavistock Square, London, W.C.1.

Papers are accepted on the understanding that they are subject to editorial revision, and that they are contributed to this Journal only. The Editors cannot undertake correspondence about papers returned as unsuitable on the advice of members of the Editorial Board.

Adequate reference should be made to previous work on the subject of the paper. A summary of observations and conclusions must be given.

Papers sent for publication must be typewritten on one side of the paper only, with double spacing and with a margin of at least 1½ inches. Illustrations should be used sparingly. Photographs should be printed on glossy paper and should be a little larger than the size desired for reproduction. Drawings and diagrams should be done in black ink on tracing linen, Bristol board, or stout, white paper. Any lettering on these drawings should be lightly inserted in pencil. The setting up of type by the printers will be much facilitated if attention is paid to the proportions of the standard page measurements. The Editors express the hope that contributors will study the conventions of shading and design of charts employed in the Journal with a view to standardizing visual symbols for ready recognition.

References should be arranged alphabetically at the end of the paper according to the Harvard system. If the reference is to a book, the place of publication, the number of the edition, and the page number should be given. In the text the year of publication must follow the author's name, more than one paper in any one year being indicated by a small letter (a, b, c) after the date. The reference should be given as follows: author's name, initials, year of publication (in parentheses), title of periodical (in italics, abbreviated according to *World Medical Periodicals*), volume number (bold type arabic numerals), and first page number (ordinary type arabic numerals); book titles (in roman and inverted commas), followed by publisher, place of publication, etc.:

Engle, E. T., and Shelesnyak, M. C. (1934). *Hum. Biol.*, **6**, 431.
Fluhmann, C. F. (1936). *Amer. J. Obstet. Gynec.*, **31**, 573.
—— (1939). " Menstrual Disorders." W. B. Saunders, Philadelphia
Gómez, F. D. (1942). *Hoja tisiol.*, **2**, 25.

Contributors will receive one proof, but it is assumed that all but verbal corrections will have been made in the original manuscript; an allowance at the rate of ten shillings per sheet of sixteen pages is made for alterations in the proof (printer's errors excepted), and contributors will be responsible for any excess.

Twenty-five reprints of papers will, if desired, be provided free. A limited number of additional reprints at cost price can be supplied if application is made when returning proofs. An estimate of the cost will be given on application to the Publishing Manager, British Medical Association.

The Journal will be published quarterly.

Papers which have appeared in THE BRITISH JOURNAL OF PREVENTIVE AND SOCIAL MEDICINE remain the property of the Journal, and permission to republish them must be obtained from the Editors.

Applications for advertisement space should be addressed to the Advertisement Manager, British Medical Association, Tavistock Square, London, W.C.1.

NOTICE TO SUBSCRIBERS

Subscriptions (£2 2s. 0d. per annum) are payable to the British Medical Association.

Brit. J. prev. soc. Med. (1957), **11**, 51–60

STATURE AND WEIGHT OF MEN FROM ENGLAND AND WALES IN 1941

BY

E. M. B. CLEMENTS AND KATHLEEN G. PICKETT

Department of Anatomy, University of Birmingham

Variation in stature and weight from one region of Britain to another is generally taken for granted: the short Welshman, the burly Scot, and the tall Southerner are accepted national types. But until recent years, there has been little scientific evidence to confirm the reality of such differences in physique, for in the 20th century no extensive anthropometric survey was undertaken to collect the necessary information about the whole population until the second world war. The nearest approach to such a study was the work of the Industrial Health Research Board between 1929 and 1932 (Cathcart, Hughes, and Chalmers, 1935), but this was mainly concerned with the principal industrial areas of England and Scotland.

Between 1939 and 1945, however, a substantial proportion of the population was measured in the course of medical examinations before National Service and of surveys made to assess the adequacy of the diet in the civilian population in war-time conditions. The data so obtained have already formed the basis of a number of reports, including the results of a survey of young adult males measured in 1939 (Martin, 1949), a survey of civilians measured in 1943 (Kemsley, 1950), and a survey of Royal Air Force personnel (Morant, 1949–50).

The present study presents an analysis of the stature and weight of men from England and Wales measured in 1941 by medical boards of the Ministry of Labour and National Service before call up for the services. It shows, in conjunction with the previously published reports relating to Scotsmen measured at the same time (Clements and Pickett, 1952, 1954), the extent of regional variation in recent years. The relationship of these measurements to each other, and to age and socio-economic status, also extends the study of Scotsmen to the whole country.

MATERIAL

Details of the material, methods of sampling, and information abstracted from the records have been given in a previous report (Clements and Pickett, 1952). The same system has been followed in the present survey.

Men on whom the survey is based were examined at ninety medical boards in England and eight in Wales. The towns in which the boards were held have been grouped into twelve regions in England and two in Wales, in accordance with the Registrar-General's geographical classification.

By means of a sampling fraction based on the total number of examinations made in the period between January 1 and March 31, 1941, a sample of approximately 200 was taken from the records available for each medical board. In all 22,315 records were collected, 1,019 of which (4·6 per cent.), mostly men with gross disability, were rejected because information was lacking. The analysis is therefore based on the measurement of 21,296 men from England and Wales.

Men attending the boards wore only trousers while being examined.

Table I shows the distribution of the sample by age.

TABLE I

DISTRIBUTION BY AGE

Age (yrs)	17	18	19	20–24	25–29	30–34	35–39	40–44	45–49
Number	62	583	5,123	4,834	2,481	2,220	5,922	45	26
Percentage	0·29	2·74	24·06	22·70	11·65	10·42	27·81	0·21	0·12

Table II shows the distribution by social class. The social classes correspond to the Registrar-General's classification of occupations (General Register Office, 1951). Students have been placed in Class 2. Classes 1 and 2 have been combined because the number of men in Class 1 is small, and Classes 4 and 5 have been combined because they were found to be very similar.

TABLE II
DISTRIBUTION BY SOCIAL CLASS

Social Class	Data (1941)			Census (1951)	British Association Anthropometric Committee (1883)
	England and Wales		Scotland		
	Number	Percentage	Percentage	Percentage	Percentage
1/2	1,676	7·9	9·2	18·3	43·2
3	10,233	48·0	47·1	52·7	33·6
4/5	9,387	44·1	43·7	29·0	23·2
Total	21,296	100·0	100·0	100·0	100·0

The social class distribution of Scottish National Servicemen measured in 1941 and of the total population of England and Wales in 1951 (General Register Office, 1954) are also given in Table II. The distributions of the two groups of National Servicemen are similar, but in comparison with the 1951 population they show under-representation of Class 1/2 and over-representation of Class 4/5. An examination of the distribution of volunteers and conscripts among Scottish National Servicemen showed a significantly greater proportion of volunteers in Class 1/2 and a significantly smaller proportion in Class 4/5, and it may be that the disproportionate distribution of classes is attributable to the earlier voluntary enlistment of a relatively larger number of Class 1/2 men.

RESULTS

STATURE.—Since Scotsmen are normally distributed by stature (Clements and Pickett, 1952), the usual tests of statistical significance are appropriate. In the following discussion all differences to which attention is drawn satisfy the conventional criteria (5 per cent. level of significance).

(*a*) *Geographical Variation.*—The Figure (opposite) shows county boundaries and the Registrar-General's geographical regions of England and Wales within which stature has been examined.

It will be seen that northern England is divided into four regions:

North I Northumberland and Durham;
North II North and East Ridings of Yorkshire with Cumberland and Westmorland;
North III West Riding of Yorkshire;
North IV Cheshire and Lancashire.

The Midlands are divided into two regions:

Midland I Western counties;
Midland II Eastern counties.

The East, South-East, and South-West, regions are not sub-divided.

Greater London is considered separately.

Wales I includes the counties of South Wales and Wales II those of North Wales.

Table III gives the mean stature in regions of England, Wales, and Scotland. In England, stature appears to diminish as one moves northwards. The lowest mean stature is found in Wales and in the most northerly counties of England.

TABLE III
MEAN STATURE AND WEIGHT IN EACH GEOGRAPHICAL REGION, IN ORDER OF STATURE

Region		Number	Mean Stature (in.)	Mean Weight (lb.)
England	South-East ..	3,929	67·56±0·04	138·2±0·3
	South-West ..	1,670	67·45±0·06	137·5±0·4
	Midland II ..	1,356	67·29±0·07	135·1±0·5
	Midland I ..	1,963	67·21±0·06	135·9±0·4
	Greater London	1,920	67·15±0·06	136·7±0·4
	East	1,512	67·09±0·07	137·5±0·5
	North III ..	2,011	67·04±0·06	135·0±0·4
	North II ..	1,289	66·73±0·07	134·6±0·5
	North IV ..	2,760	66·73±0·05	132·9±0·3
	North I ..	1,113	66·51±0·08	134·3±0·5
	All	19,523	67·14±0·02	135·9±0·1
Wales	Wales II ..	628	66·60±0·11	134·7±0·7
	Wales I ..	1,145	66·53±0·08	133·2±0·5
	All	1,773	66·55±0·06	133·7±0·4
Scotland	Southern ..	409	66·96±0·13	140·6±0·9
	Northern ..	1,285	66·94±0·07	141·9±0·5
	East Central ..	895	66·78±0·08	135·8±0·6
	West Central ..	1,103	66·65±0·08	134·8±0·5
	All	3,692	66·82±0·04	138·2±0·3
Total England, Wales, and Scotland		24,988	67·05±0·02	136·2±0·1

Since the previous analysis of Scottish data (Clements and Pickett, 1952) indicated that socio-economic status is the main source of variation in adult stature, the data for England and Wales have been divided into three social class groups. An analysis of variance was used to test differences between the statures of men in the same social class who were examined at different boards in the same region. The results enabled "homogeneous regions" to be built up, that is, regions within which no significant variation in stature was found.

There was no significant regional variation in stature in Class 1/2, except in the case of one board

(Barnsley) in region North III, where the average stature was low.

In Class 3 five boards showed deviations from the prevailing stature of the region: men from the Tottenham board (Greater London) were shorter; men from Doncaster and Sheffield (North III) and Swansea and Cardiff (Wales I) were taller. No unusual circumstance was apparent to account for the Cardiff board's deviation, but Tottenham had a raised incidence of men over 30, and the other three boards had a raised incidence of men under 20 years of age.

In Class 4/5, seven boards showed statures greater than the remainder of their regions: Hounslow and Willesden (Greater London); Taunton, Bath, and Weymouth (South-West); Cambridge (East); and Barrow (North IV). Except in the case of Weymouth, where the sample was disproportionately young, there was no obvious explanation for these differences.

It seems hard to believe that the deviations from the mean stature of other boards in the same region which have just been detailed reflect a genuine geographical variation on what can only be a minute scale. No one factor seems responsible and the explanation remains obscure although there is perhaps a hint that the more urbanized or industrialized centre has an above average stature for the region.

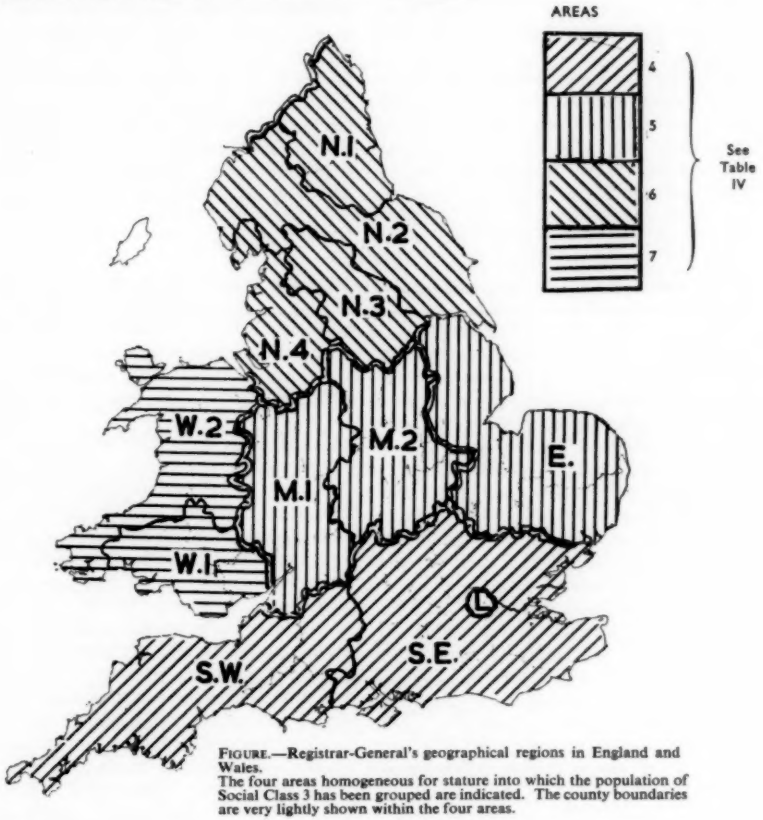

FIGURE.—Registrar-General's geographical regions in England and Wales.
The four areas homogeneous for stature into which the population of Social Class 3 has been grouped are indicated. The county boundaries are very lightly shown within the four areas.

REGIONS

S.W. South-west (Cornwall, Devon, Dorset, Somerset, Wilts.)
S.E. South-east (Hants., Sussex, Surrey, Kent, Oxon., Berks., Bucks., Herts, Beds, Essex.)
L. Greater London
E. East (Cambs., Lincs., Norfolk, Suffolk)
M.1. Midlands I (Salop, Staffs., Glouces., Herefs., Worc.)
M.2. Midlands II (Derbys., Notts., Leics., Warwicks., Northants.)
N.1. North I (Northumb., Dur.)
N.2. North II (Cumb., Westmorland, N. Rid., E. Rid.)
N.3. North III (W. Rid.)
N.4. North IV (Ches., Lancs.)
W.1. Wales I (Carmarth., Breck., Glam., Mon.)
W.2. Wales II (Pembs., Radnor, Cards., Mont., Merioneth., Caernarvon, Anglesey, Flint, Denbighs.)

The data from the unusual samples have been excluded from the computation of the statistics for the various regions. Within each social class group, these statistics have been combined to form areas over which an analysis of variance indicated no significant deviation in stature; Table IV shows the ten groups obtained, three for Classes 1/2 and 4/5, and four for Class 3 (Figure).

Although the combination of regions is not identical for each social class, a similar pattern emerges: decreasing stature from south to north, with Wales always in the group with the shortest mean stature. The two Midland regions are the most stable; for there is little difference between them or their constituent boards, and they are always found in the same group as Greater London and the Eastern region.

A very similar distribution was found by Cathcart, Hughes, and Chalmers (1935) for men measured in England and Scotland between 1929 and 1932, and by Martin (1949), whose figures refer to men in England, Scotland, and Wales between the ages of 20 and 21, measured in 1939. Martin's data are not generally comparable with those of the present survey, as the age range and the regions into which the men are grouped are very different. Both surveys, however, confirm that the average stature of Englishmen in the south is greater than that of men both in the north of England and in Scotland. It seems to be clear that the factors leading to variation in stature from one part of the country to another affect all sections of the community to a similar extent.

The relative position of men living on the borders of Scotland is perhaps the most surprising. Beddoe (1870), who made the first regional study of the stature and weight of British men, found the inhabitants of the Border country to be the tallest in his sample, which was due, he believed, to a "strong Scandinavian element" among them. Some of his shortest subjects came from the southern counties, but his average for Lancashire was "as low or lower than the rest of England", and similar conclusions were drawn by the British Association Anthropometric Committee of 1883. Table II, however, shows that the social class distribution

TABLE IV

MEAN STATURE (IN.) OF AREAS IN ENGLAND AND WALES SHOWING NO SIGNIFICANT VARIATION IN STATURE

Social Class								
1/2			3			4/5		
Area	Region	Mean Stature and Standard Deviation	Area	Region	Mean Stature and Standard Deviation	Area	Region	Mean Stature and Standard Deviation
1	South-East	68·52 ± 0·11 S.D.: 2·64	4	South-East	67·66 ± 0·05 S.D.: 2·50	8	South-East	67·22 ± 0·07 S.D.: 2·63
	South-West			South-West				
2	Greater London	67·91 ± 0·09 S.D.: 2·57	5	Greater London, without Tottenham	67·44 ± 0·04 S.D.: 2·58	9	South-West, without Bath, Taunton and Weymouth	66·68 ± 0·04 S.D.: 2·68
	East			East			Greater London, without Hounslow and Willesden	
	Midlands I			Midlands I			East, without Cambridge	
	Midlands II			Midlands II			Midlands I	
	North II		6	North I	66·96 ± 0·05 S.D.: 2·64		Midlands II	
	North III, without Barnsley			North II			North III	
	North IV			North III, without Doncaster and Sheffield				
				North IV				
3	North I	67·02 ± 0·17 S.D.: 2·64	7	Wales I, without Cardiff and Swansea	66·55 ± 0·11 S.D.: 2·58	10	North I	66·30 ± 0·05 S.D.: 2·68
	Wales I			Wales II			North II	
	Wales II						North IV, without Barrow	
							Wales I	
							Wales II	

N.B. Each group of homogeneous regions has been numbered, and the areas formed are identified in the text by these numbers.

in at least the latter sample was not representative and it seems likely that, in both reports, differences in stature attributed to regional variation are the result of the variation from place to place in the proportion of men of differing socio-economic environment. Stature differences between social class groups appear to have been greater at that time than they are to-day, and their uneven representation may be one reason for the discrepancies.

Migration since the time of the earlier surveys may also have had its effect. The figures given by Martin (1949) show that immigrants into a region, defined as men whose birthplace differs from the town in which they registered for National Service, are almost invariably taller than natives, and that migration to the southern counties has been at a particularly high level. Immigrants may even form over a third of the population in parts of the south-east and south-west of England, while in the north they may reach only 15 per cent. or less of the total. In Scotland the figure for immigrants is 3 per cent. and in Wales 6 per cent. These proportions, as Martin admits, may be over-estimates, but, although some of the men may have moved in childhood, the figures are sufficient to show a very marked trend from north to south.

The mean of 5′ 7″ for the whole country, obtained when all the social class groups are pooled, agrees with Beddoe's estimate made in 1870 of between 5′ 6″ and 5′ 7″; it is however a little shorter than that of 5′ 7½″ given by the British Association Anthropometric Committee in 1883. Beddoe's sample includes a much higher proportion of working men and middle-class men than does the survey of 1883, and his estimate is probably closer to the national mean at that time.

Some small part of regional variation in stature may be of racial origin, but the data suggest that only in Wales and some areas of Scotland has isolation been sufficient for racial characteristics to survive into the 20th century. The Celts, a relatively short-statured race, who moved into Wales, have remained until recent years comparatively undiluted by the taller races who became the main colonizers of Britain. Elsewhere, apart from islanders in the far north, where the descendants of invading Danes of the 8th century are taller than the mainland Scots, racial intermarriage has resulted very largely in a loss of identity. It is probable that other factors, such as migration, encouraged by increasing industrialization, have made an important contribution to the pattern found to-day.

(*b*) *Socio-Economic Status.*—In not one of the three groups (all England and Wales) for the composite Class 1/2 does the mean stature of men of Class 1 differ significantly from that of Class 2. However, in a similar examination of the Scottish data a significant difference (1·2″) arose between the mean statures in these two classes (Clements and Pickett, 1952), and it is possible that a high degree of variability occurs within Class 1 itself.

Pooled values for the three social class groups (Table V) show a decrease of 0·7″ from Class 1/2 to Class 3, and a decrease of 0·6″ from Class 3 to Class 4/5. In 1883, the British Association Anthropometric Committee found a difference of almost 2″ between "professional and commercial classes" and "artisan classes living in towns".

TABLE V

MEAN STATURE (IN.), BY AGE AND SOCIAL CLASS

Age Group (yrs)	Social Class			
	1/2	3	4/5	All
17–19	68·48 ± 0·14	67·53 ± 0·05	67·07 ± 0·06	67·42 ± 0·04
20–29	68·05 ± 0·11	67·42 ± 0·04	66·89 ± 0·05	67·27 ± 0·03
30–39	67·61 ± 0·10	66·89 ± 0·05	66·41 ± 0·04	66·68 ± 0·03
40–49	68·57 ± 0·76	66·95 ± 0·49	65·94 ± 0·49	66·78 ± 0·32
All	67·97 ± 0·06	67·30 ± 0·03	66·70 ± 0·03	67·09 ± 0·02

It is evident from Table IV that, although there is a decrease from Class 1/2 to 3 and from Class 3 to 4/5 in the stature of men from similarly located areas, a comparison of samples from widely separated areas may show Classes 3 or 4/5 to have a greater mean stature than Class 1/2.

Thus, although the mean stature of Class 1/2 men in the two groups which include all the English regions except North I (Areas 1 and 2 in Table IV), is greater than that of men in any other group, that of Class 1/2 men in the group which includes Wales and North I (Area 3) is less than that of men in Class 3 in most southern and midland regions (Areas 4 and 5) and of men in Class 4/5 in the South-east (Area 8). There is no significant difference between the mean stature of Class 3 men in the Midlands, Eastern region, and Greater London and that of Class 4/5 men in the South-east (Areas 5 and 8). The latter are taller than those of Class 3 in the Welsh regions (Area 7).

(*c*) *Industry and Occupation.*—The variation of stature was studied in different industries, and in different occupations in the same industry, within each social class group, and within one group where it was known that there was overall homogeneity of stature (Areas in Table IV).

For the purpose of this analysis, to increase the size of the samples and so to provide sub-samples of reasonable size, the discrepant boards have been included and placed with a regional group with which they combine homogeneously. As has been made clear earlier, this should not be taken to imply a geographical association between these areas.

These boards have been included in regional areas as follows:

Class	Board	Regional Area
1/2	Barnsley	3
3	Doncaster and Sheffield	4
	Cardiff and Swansea	5
	Tottenham	7
4/5	Bath, Taunton, Weymouth, Hounslow, Willesden, Cambridge and Barrow	8

Within each social class, the groups containing the greatest numbers were chosen. For each test a covariance analysis was made in which age was the independent variable, to ensure that age differences between the samples were taken into account when making the comparison.

In Class 1/2 an analysis was made of the stature of men from Area 2 (Table IV), and no significant differences among men in the various industries represented were found. Differences at the 5 per cent. level of significance occurred among men in professional occupations.

Men from Area 5 (Table IV) represented Class 3, and here there were marked differences according to industry. This appeared to be due to two groups, one unusually tall and the other unusually short. The first included 463 clerks and typists with an average height of 68·2″ (compared with 67·4″ for this group as a whole). The second group included men in personal service and in this group it was found on further examination, that the low average was mainly due to a sample of 36 hairdressers (65·6″) and twelve chefs (66·4″).

An analysis of men in the same class working in the transport and engineering industries indicated no differences related to occupation.

Substantial differences according to industry were again observed among men of Class 4/5 from Area 9 (Table IV). The greatest deviation occurred in a sample of 27 oilers and greasers whose mean stature was 65·9″ compared with 66·7″ for the group as a whole.

Men in Class 4/5 in the engineering industry showed no occupational variation in stature. There were marked differences in the transport industry, however, where the mean stature of 36 locomotive firemen was 68·4″ compared with 67·1″ for all men in the industry, but these men may perhaps be wrongly classified as semi-skilled workers. Other groups of men with divergent mean heights include 23 messengers (65·9″), 46 porters (66·3″), and 26 dock labourers (66·2″).

Variation of stature within industries and occupations appears in the main to be very slight. A few divergent groups occur, but in most cases these are probably the result of sampling irregularities. The high mean stature of clerks and typists compared with that of others of their economic group, however, has practical significance, for example in the design of office furniture.

(*d*) *Age.*—Notwithstanding the factors of geographical and socio-economic variation which may affect adult stature, its relationship with age, as demonstrated by regression, does not differ significantly between regional groups or between classes. Table V gives the mean stature for four age groups covered by the sample, and shows that the stature of younger groups is greater than that of older ones. (The trend is less clear because the 40 to 49-year group is a relatively small sample consisting entirely of volunteers). Analysis of the data relating to Scotsmen (Clements and Pickett, 1952) gave the same result and confirms similar findings by Kemsley (1950) and Morant (1950). The regression coefficient of stature on age for the pooled data is $-0{\cdot}048 \pm 0{\cdot}002$″/yrs.

Differences between social class means are little affected by standardization for age.

There is a significant difference between regressions on age within the four age groups; and, as the age distribution of the samples for England and Wales and for Scotland differ, this factor probably explains the difference found between the regressions. This difference between age groups may be mainly due to the 17 to 19-year age group, in which stature appears still to be increasing. The data for England and Wales confirm the earlier observation on Scottish data that in all social classes maximum adult stature is reached during the 19th year.

WEIGHT.—Body weight is not normally distributed, but shows an excess around the mean, or kurtosis, and an excess at the upper end of the distribution, or skewness (Clements and Pickett, 1954). Analysis of the Scottish National Service data indicated that log weight, on the other hand, may be considered to be normally distributed for all practical purposes. Individual weights have therefore been converted to their logarithms to the base 10. Log weight has been used throughout the analysis, and conversion to weight on the arithmetic scale for the purposes of Tables III and VII has been made by means of the equation:

Observed mean weight (lb.) = 1·045 mean log weight (converted to lb.)—5·098. (Clements and Pickett, 1954).

(*a*) *Geographical Variation.*—Table III gives mean weights for regions of England, Wales, and Scotland, pooling all social classes. In the main, the trend for weight in England and Wales follows that for stature, though men in the Eastern region tend to be rather heavy for their stature, whereas men of the Midland II region are rather light.

The heaviest mean weights in Britain are found in the northern and southern regions of Scotland, although stature here is close to the mean of the whole country. Both Beddoe and the British Association Anthropometric Committee at the end of the 19th century found that the Scots were heavier than the English or Welsh, and Martin (1949) also reported that young men in the Highlands, measured in 1939, were heavier than those from all other regions. When the mean weights of all regions are adjusted to constant stature, those of Scotsmen remain greater in every age group than those of the English and Welsh.

The mean log weight of men in each group in which there is no significant variation in stature (Areas in Table IV) is given in Table VI.

TABLE VI

MEAN LOG WEIGHT OF AREAS IN ENGLAND AND WALES SHOWING NO SIGNIFICANT VARIATION IN STATURE

Area (see Table IV)	Social Class		
	1/2	3	4/5
1	2·1509 ± 0·0024		
2	2·1408 ± 0·0020		
3	2·1387 ± 0·0038		
4		2·1358 ± 0·0010	
5		2·1324 ± 0·0009	
6		2·1248 ± 0·0010	
7		2·1236 ± 0·0025	
8			2·1346 ± 0·0013
9			2·1269 ± 0·0009
10			2·1201 ± 0·0009

The regions making up these groups do not show a corresponding homogeneity when weight is considered. Regions within Class 1/2 alone are the same in every respect for both measurements. In Class 3, the weights of men in seven boards differ significantly from those of the remainder of their regions. These boards are Lincoln (East), Canterbury and Brighton (South-east), and Barnsley (North III), with greater weights, and Sunderland (North I), Blackburn and Salford (North IV), with smaller weights. In Class 4/5, eight boards show a similar variation. Men from High Wycombe and Kingston (South-east), Plymouth (South-west), Preston (North IV), and Aberystwyth (Wales II) are heavier, and men from Sunderland (North I), and Bradford and Halifax (North III) are lighter than the others of their regions.

The reason for these deviations is obscure, but weight is generally regarded as being more sensitive to external conditions than stature.

(*b*) *Socio-Economic Status.*—Observed mean weights of men in the three social class groups have been computed using the total number of records available (21,296) and are given in Table VII. There is a decrease of over 4 lb. from the mean weight of men in Class 1/2 to that of men in Class 3, but of only 1 lb. from Class 3 to Class 4/5. After standardization to constant stature, these differences are much reduced, the mean weight of men in Class 4/5 being only 1lb. greater than that of men in Class 3, and under 1lb. less than that of men in Class 1/2.

Weight differences between regional groups, although again decreasing from Class 1/2 to 3 and from Class 3 to 4/5, when similarly located groups are compared, are not always significant. The mean log weights of Class 1/2 men in all regions except the South-east and South-west (Areas 2 and 3 in Table IV) are not significantly different. Those of Class 1/2 men in the Welsh and North I regions (Area 3), Class 3 men in all English regions excepting the Northern (Areas 4 and 5), and Class 4/5 men in the South-east (Area 8) also do not differ. The mean log weights of Class 3 men in the Welsh and Northern regions (Areas 6 and 7) do not differ significantly from those of Class 4/5 men in the English regions, excluding the South-east and North I, II, and IV (Area 9).

There are no significant differences between the mean log weights of men in Class 1 and Class 2 in any part of England and Wales.

(*c*) *Stature.*—The relationship between log weight and stature may be considered linear for all practical purposes (Kemsley, 1952; Clements and Pickett,

1954). Regression coefficients relating the two measurements in each regional group homogeneous for stature have been calculated separately, and do not differ significantly within the three social class groups. When the regional group means are standardized for stature, they also do not differ except in Class 4/5.

There are significant differences between the regressions within each social class, and between the class means after standardization, non-homogeneity being introduced by Class 1/2. The class regressions cannot therefore be pooled. Regression coefficients for each social class group are:

Social Class 1/2 0·0107 ± 0·0005 log lb./yrs
Social Class 3 0·0119 ± 0·0002 log lb./yrs
Social Class 4/5 0·0119 ± 0·0002 log lb./yrs

The correlation coefficient relating weight and stature in men from England and Wales is 0·57, compared with 0·59 in men from Scotland.

(*d*) *Age.*—Regression coefficients relating log weight to age within each regional group homogeneous for stature are, except in Class 3, consistent only between these groups. This may be in part because there has been no standardization of stature.

Except in Class 1/2, a quadratic relationship between the two is preferable, for there is significant curvilinearity in all other groups. This means that in the age range covered by the data, an equation including the square of age will give a better result than one based on age alone. The best fit is given by the following regression equations:

Social Class	Area	Weight (log lb.) =
1/2	1	0·000352 Age (yrs) + 2·141
	2	0·000905 Age (yrs) + 2·117
	3	0·00116 Age (yrs) + 2·106
3		0·00857 Age (yrs) − 0·000147 Age² + 2·015
4/5	8	0·00760 Age (yrs) − 0·000138 Age² + 2·039
	9	0·00206 Age (yrs) − 0·0000323 Age² + 2·097
	10	0·00527 Age (yrs) − 0·0000848 Age² + 2·044

* Areas are those of Table IV modified as explained in section "Industry and Occupation".

The unstandardized mean weights of men in each age and social class group (Table VII) illustrate the differences in the weight-age relationship which arise between the classes. The lightest mean weight is always found in the 17 to 19-year age group, but, while there is a relatively constant increase within Class 1/2, there is little or no difference between the weights of men of 20 to 29 years and of 30 to 39 years within Classes 3 and 4/5 (ignoring the 40 to 49-year age group which is small and unrepresentative).

Standardization to constant stature, however, shows that, in all classes, men in older age groups are heavier in relation to stature than younger men. Table VII (opposite) gives standardized mean weights for pooled classes; and the standardized means for separate classes are very similar. Comparable figures are given by Kemsley (1950) for the civilian population in 1943. Both surveys indicate that increasing weight with age is much less marked than in 1883 (British Association, 1883), or even between 1929 and 1932 (Cathcart, and others 1935). Whether this is a persistent trend or is attributable to wartime dietary conditions, which might be expected to have the least effect on men of favourable socio-economic background, is not clear.

(*e*) *Stature and Age.*—A multivariate analysis shows that, after standardization for both stature and age, including the square of age, log weight differences between the regional groups homogeneous for stature disappear in Classes 1/2 and 3, and the regressions relating to men in these classes may be pooled. This gives the following relationship for the pooled classes and for the regional groups of Class 4/5:

Social Class	Area	Weight (log lb.) =
1/2		0·0112 Stature (in.) + 0·00363 Age (yrs) − 0·0000424 Age² + 1·317
3		0·0122 Stature (in.) + 0·00850 Age (yrs) − 0·000136 Age² + 1·192
4/5	8	0·0116 Stature (in.) + 0·00638 Age (yrs) − 0·000106 Age² + 1·264
	9	0·0126 Stature (in.) + 0·00298 Age (yrs) −0·0000385 Age² + 1·236
	10	0·0121 Stature (in.) + 0·00847 Age (yrs) − 0·000134 Age² + 1·192

* Areas are those of Table IV modified as explained in Section "Industry and Occupation".

TABLE VII

MEAN WEIGHTS (LB.), BY AGE AND SOCIAL CLASS GROUPS

Age Group (yrs)	Social Class								Standardized Means for All Social Classes
	1/2		3		4/5		All		
17–19	138·1	(382)	133·8	(3,231)	133·0	(2,155)	133·8	(5,768)	132·6
20–29	139·4	(595)	137·1	(4,005)	135·3	(2,715)	136·6	(7,315)	135·9
30–39	142·4	(687)	137·1	(2,969)	135·5	(4,486)	136·7	(8,142)	138·2
40–49	146·9	(12)	148·4	(28)	136·8	(31)	143·1	(71)	144·3
All	140·4	(1,676)	136·1	(10,233)	134·9	(9,387)	135·9	(21,296)	—
Standardized Means for All Age Groups	137·1		135·3		136·4		—		—

The numbers in each sample are shown in brackets.

DISCUSSION

The results reported in earlier sections of the paper provide information on the distribution of physique in the male population. They confirm the existence of national types and show in greater detail the pattern linked to geographical factors which underlies the large normal variation of everyday experience. Stature becomes lower on passing from the South to the North of England or into Wales, whereas weight is greater in Scotland. Socio-economic factors apparently account for most differences in industry and occupation, and on the whole, differences within the same class are small. A number of questions, however, remain unanswered. Thus it is not known whether the pattern of increase in weight with age remains as it was found to exist in war-time conditions, or has now reverted to the increasing trend evident earlier, or perhaps follows some new pattern. We have suggested that the unusual samples which differed from similar samples in the same region may be in some way connected with urbanization, but whether the essential factor is one of migration, genetic origin, or simply one of occupation remains unanswered.

The statistics may serve in addition as standards to which other material may be compared. Taken together, the measurements of stature and weight permit a classification of physique to be made in a quantitative way which will estimate the frequency of occurrence of physical types in the population. Up-to-date estimates have been derived from the material relating to Scotland (Clements, 1955), and the smallness of the difference in the statistics for Scotland, when compared with those for England and Wales makes it clear that for practical purposes these estimates may be applied to the country as a whole.

The statistics may also be applied to the construction of a sizing system for clothing or in the design of equipment. In this connexion, where a product is to be used by the population as a whole, for practical purposes the wide range normally found in the size of body dimensions makes it unnecessary to allow specifically for local variation, because the difference between one locality or section of the community and another may be taken into account by using a different proportion of each size depending on the characteristics of the particular group they are expected to serve.

SUMMARY

(1) The mean stature and weight of men between the ages of 17 and 50 years measured in 1941 by medical boards of the Ministry of Labour and National Service are given for each of the Registrar-General's geographical regions in England, Scotland, and Wales.

(2) Significant differences are shown to exist between men of different socio-economic status, and of the same socio-economic status in different areas of the country.

(3) The stature of men of the same socio-economic status and geographical location, working in different industries and in different occupations within the same industry, are examined.

(4) It is confirmed that maximum stature is reached during the 19th year. A continuous increase in weight with age occurs only in men in professional and managerial occupations.

(5) The relationships between stature and age, stature and weight, and weight and age are discussed and regression coefficients are given.

We gratefully acknowledge the receipt of a grant from the Medical Research Council and the assistance given by the Ministry of Labour and National Service. We should also like to thank Prof. Sir Solly Zuckerman, C.B., F.R.S., Dr. M. J. R. Healy and Mr. W. F. F. Kemsley for advice in many helpful discussions.

REFERENCES

Beddoe, J. (1870). "On the Stature and Bulk of Man in the British Isles." Asher, London. Reprinted from *Mem. anthrop. Soc. Lond.*, 1867–9, **3**, 384.

British Association (1883). Final Report of the Anthropometric Committee. *In* "Report of the 53rd meeting of the British Association for the Advancement of Science, Southport, 1883", p. 253.

Cathcart, E. P., Hughes, D. E. R., and Chalmers, J. G. (1935). "Physique of Man in Industry". M.R.C. Industrial Health Research Board, Rep. 71. H.M.S.O., London.

Clements, E. M. B. (1955). *Voeding*, **16**, 51 (3rd International Congress of Nutrition).

—— and Picket, K. G. (1952). *British Journal of Social Medicine*, **6**, 245.

—— (1954). *British Journal of Preventive and Social Medicine*, **8**, 99.

General Register Office (1951). "Classification of Occupations, 1950". H.M.S.O., London.

—— (1954). "Census 1951, England and Wales". H.M.S.O., London.

Kemsley, W. F. F. (1950). *Ann. Eugen* (*Camb.*), **15**, 161.

—— (1952). *Ibid.*, **16**, 316.

Martin, W. J. (1949). "The Physique of Young Adult Males". M.R.C. Memo. No. 20. H.M.S.O., London.

Morant, G. M. (1950). *Proc. roy. Soc. B.*, **137**, 443.

—— (1949–50). Flying Personnel Research Committee. Reports 711, 711(a), 711(b), 711(c), 711(d).

Brit. J. prev. soc. Med. (1957), **11**, 61–72

RANDOM SELECTION IN THE THERAPEUTIC TRIAL

BY

R. F. WRIGHTON

Department of Medical Statistics, University of Birmingham

I. INTRODUCTION

This paper completes the formal part of an inquiry into the credentials of the statistical method as related to the evaluation of therapeutic measures. I have summarized the general conclusions of this inquiry in an earlier communication (Wrighton, 1955). What follows is an attempt to elucidate concepts which, whilst of less immediate relevance than those I have dealt with earlier (Wrighton, 1953), are, in the final analysis, of equally vital importance.

In employing concepts not peculiar to his own discipline, the biologist can often draw on a vocabulary refined and tested in the laboratory so as to be for all practical purposes quite unequivocal. This is regrettably not so in the case of statistics. The relevant aspect of physical reality is seemingly no-one's professional responsibility, unless it be the mathematician's. But mathematicians have never approached unanimity on the subject and, while only too ready to suggest technical procedures owing their origins to unstated and dubious assumptions, have in the main seemed unprepared to admit that fundamental inquiry is necessary or profitable.

The laboratory situation we need to examine in this context is properly remote from the practical problems which are our ultimate concern. We have an urn containing unknown numbers of balls of various specifications.* We may then ask: first, what precise meaning can we have in requiring that a number of balls should be drawn from the urn *at random*? Secondly, assuming the first question answered, what kind of information about the contents of the urn can we legitimately derive from a sample thus drawn? The second is the question I have already attempted to answer. Semantic confusion is there associated with words such as *significance*, *confidence*, and *likelihood*, and the notion of *inverse probability*. The first is the question I am concerned to answer here. As I have suggested above, it is of more than technical interest. Indeed it must, strictly speaking, be answered before the other can be approached with complete propriety. The semantic issues here relate to the notions of *chance event*, *random procedure*, and *direct probability*.

The method of random selection in common use requires the numbering of the balls in the urn, and uses published lists of random digits to effect the withdrawal. There are two types of objections we can make to such a procedure. We can object that the use of an established table of random numbers is illegitimate, however the table has been compiled, or we can object specifically to the mode of structure of the table, viewed in itself as a dynamic randomizing device. In considering such objections, I must slightly anticipate discussion in the body of the paper and use loosely terms whose meanings I intend later to make more precise.

With any randomizing procedure we anticipate that successive results will be discordant, yielding, on appropriate tabulation, a chaotic array in which no pattern can be detected. The compiler of a table of random digits can be judged *after the event* only by his success in producing such an appearance. To do so, however, he may or may not invoke what we can refer to as a *disorderly procedure*. In the fact of the existence of this choice lies an important objection to the method of random digits. Suppose, for example, we have equal numbers of black and white counters filling the squares of a chess-board in a systematic arrangement. We may use the disorderly procedure of shuffling in order to redistribute the counters amongst the squares; and we can suppose, if our shuffling is vigorous, that *almost certainly*, no new pattern will appear. We may, however, proceed quite differently in order to obtain the same type of

* I have given the specifications relevant to the simplest type of therapeutic trial in the second of the communications cited above.

result, for we may, though with great labour, systematically enumerate the possible arrangements of counters on the board and select from amongst these one which passes all of a number of pre-assigned and systematically compiled tests for freedom from pattern.* We are led, then, to distinguish sharply between the concept of a static chaos and the concept of a chaotic procedure. The latter may be sufficient, but it is not necessary for the production of the former. With our present preoccupation we are concerned only with chaotic procedures. If we use a static chaos we cannot hope to ensure that the appearance of a particular sample is in the strictest sense unanticipated.

A further aspect of the failure of the static random digit table to fulfill our need may be briefly mentioned. Suppose, on the one hand, the compiler of such a table uses a disorderly procedure to effect his end, and produces, against long odds, a table with markedly systematic features. If he does not wish to pass this misfortune on to the user, he will either suppress the table or give warning of the regularities present.† In either case he admits that the calculus of probability is not strictly applicable when the table is used. If, on the other hand, no undesirable regularities appear, the compiler is no better off; if subsequent calculations are to be strictly valid, he must bear in mind the whole range of possible outcomes of the random procedure and must examine his conscience to discover whether any result conceivable could have caused his behaviour to vary. We must conclude, therefore, that the use of random number tables is quite inconsistent with the end we have in view.

We consider now objections to currently suggested methods of compiling random number tables, viewed as randomizing devices in their own right. Some methods we must regard as *a fortiori* invalid on grounds already discussed. These include any method which uses printed or otherwise predetermined data, such as end digits in mathematical tables and census reports. There remain *ad hoc* and infinitely protractable procedures. One such method takes as random numbers the middle digits of arbitrarily chosen large numbers multiplied together on a modern digital computer. This method, and others like it, we must reject out of hand, since our ideas can be little clarified in replacing the word *random* by *arbitrarily chosen*. Another method utilizes so-called stochastic phenomena met with in physics. This too we must reject, since in principle at least we must attend to the possibility of fresh experimental inquiry vitiating any assumption we make. There remain methods (such as that of Kendall and Babington Smith) which make direct use of apparatus similar to that used in familiar games of chance. We shall see in what follows that it is possible to formulate a satisfactory approach on this basis; but to do so we shall have to make quite clear that a game of chance is and, furthermore, make a vital emendation of the procedure which takes over unqualified the results which flow directly from a chance apparatus used therein. In doing this we shall in effect be providing new supports to the foundations of a part of probability theory which is at least of great historical importance.

Since we are not concerned here in developing a theory descriptive in aim, we shall conveniently adopt a genetic approach, finding accordingly that in laying the foundations for a formal development we can at the outset provide ourselves with the means of recognizing those situations in the real world to which the system must be regarded as applicable. In doing so we shall avoid the paradoxes encountered by those who seek, as it were after the event, to make the domain of application of probability theory commensurate with a wide and vaguely defined domain of indeterminate behaviour. It follows that, while we shall make claim to associate an unequivocal meaning with the term probability in the limited context of situations in which stochastic events are purposively evolved, we shall not pretend that this directly clarifies the sense in which the word is or should be used, for example, in physics, in genetics, in general statistics, or in logic.

II. The Concept of Fortuity

The following is the problem which we seek to resolve: given N distinguishable objects, to derive an unambiguous procedure by which we can select n objects from amongst them, so that, immediately before the act of selection, every such combination of objects may be said to possess the same opportunity of being withdrawn. We refer to this as the lottery problem.

We can effect a casual resolution by matching the objects concerned with numbered tickets, shuffling these together more or less thoroughly and calling upon a disinterested person to withdraw from

* What amounts to a specification of linear patterns in a similar context is given by Popper (1935).

† Kendall (1945, p. 196) remarks:

"Thus it is to be expected that in a table of Random Sampling Numbers there will occur patches which are not suitable for use by themselves. The unusual must be given a chance of occurring in its due proportion, however small."

Kendall and Babington Smith (1938) attempted to deal with this problem by indicating the portions of their Table (5 thousands out of 100) which it would be better to avoid in sampling experiments requiring fewer than 1,000 digits.

amongst them a batch of the requisite size. In the simplest situation, where $N=2$ and $n=1$, we may adopt a more clearly defined procedure. We may toss a coin and, following a pre-assigned convention, withdraw the object labelled 1 or the object labelled 2 according as the coin lands with heads or tails uppermost. Here we may reasonably allow a potential beneficiary to operate the mechanism employed. Since the *chance* events concerned derive from a procedure for the following of which clear-cut instructions, seemingly admitting of no wilful misinterpretation, can be given, we approach a situation in which impartiality is strictly ensured and in which the notion of *chance* acquires objectivity. It is characteristic of such procedures that it is possible to verify their efficacy quantitatively. The validity of the coin-tossing procedure in particular is attested to by our experience that, with ordinary precautions and an approximately symmetrical coin, no operator whatever his efforts, can succeed in deflecting the proportion of heads observed in a large number of trials from the neighbourhood of $\frac{1}{2}$. With this sort of situation in mind we can proceed to develop a general approach to the problem we have under consideration.

In discussing the generalized gaming device, we must distinguish sharply between the physical apparatus employed—for example a coin, the medium in which it is tossed, and the surface upon which it lands—and the instructions, in practice often ill-defined, which dictate the actions of an operator as he manipulates the apparatus. More precisely, in such situations, we have to bear in mind three entities—the apparatus employed, the operator using the apparatus, and the instructions issued to the operator. We shall require first to examine the limitations imposed by unavoidable imperfections in the apparatus employed, and secondly to establish the minimum precise formulation necessary to make the operator's instructions adequately explicit. We must in particular ensure that instructions for the evolution of stochastic sequences shall contain no word—for example *shuffle*, *haphazard*, *erratic*, *nonpurposive*—whose inclusion would imply that we allow the responsibility for the impartiality hoped for in the operator's behaviour to depend in some way on his own interpretation of the purpose behind the procedure. To succeed in this will be to find a definition of the term *probability* which gives the word an important part of its connotation in the classical mathematical theory, but which anticipates charges of circularity and equivocality.

Let us ask at this point what it is that gaming devices of seemingly proven adequacy have in common. It will be seen that:

(*a*) A routine action is prescribed which may have one of a specified number of distinguishable outcomes, each associated with one facet of a symmetrical apparatus.

(*b*) The instructions given to the operator of the device are such that, *whoever he is*, it lies quite beyond his powers of sensory discrimination to predetermine any particular one of these outcomes.

We examine these properties in a chance mechanism which is conveniently simple.

Consider a teetotum formed from a disk transfixed at its centre by a style and divided into r equal sections, so that, when the operator spins it, any one of r outcomes may result according to what point on the circumference of the disk comes to rest in contact with the surface on which it is spun. We may plot the results of successive spins upon a straight line which is marked off at intervals of 2π units from an origin representing the fixed point from which the top is set in motion. We allow the operator to be any human being capable of carrying out our instructions. We instruct him to twist the top so that it spins completely round at least n times.

Let us consider the simplest case, $r=2$, the disk being divided by a diameter into two sections, one coloured black, the other white. We may distinguish three cases:

(*i*) *n small.*—The operator may acquire such skill that he can predetermine whether the outcome is *black* or *white*. We may refer to sequences of labelled events derived in this way as *determinate*.

(*ii*) *n intermediate.*—The operator is still able to exercise his choice but not with invariable success; fluctuations in the labelled sequence of events occur and are attributable to intrinsic factors, such as fatigue, or to seemingly trivial external circumstances distracting the operator or subverting the train of events he has set in motion. Resulting sequences we may refer to as statistical or *indeterminate*. With this type of instruction we can devise a game of skill between operators of comparable aptitude.

(*iii*) *n large.*—The operator is unable to influence the outcome of a trial in any way. Resulting sequences we may refer to as *fortuitous*. With this type of instruction we can devise a true game of chance.

Following this distinction, we may regard a line upon which the outcomes of individual trials are

mapped, as divided into three domains: of *determinacy*, *indeterminacy*, and *fortuity*. Vagueness in the definition of the boundaries of these domains need not trouble us. In serious application we shall wish to deal exclusively with sequences of events we can regard beyond all question as either determinate or fortuitous. Statistical series, as defined above, will not submit themselves to formal manipulation.

The principal assumption we shall need to make is that domains of fortuity exist, that is to say, that fortuitous behaviour exists, as distinct from what we have called indeterminate behaviour. Clearly a direct method of demonstrating this is inconceivable. Any evidence we may adduce in its favour must be incomplete insofar as it will relate to protracted sequences of trials. A *statistical* theory is not invalidated if occasional events within a sequence supposedly obeying probabilistic laws are, according to a particular criterion, determinate; it may merely, and perhaps to an insignificantly small extent, be inadequately descriptive. But the lottery problem can admit of no resolution at all if anomalously determinate events are allowed to intrude into the procedure of selection used; for here a repetition of the procedure, within which an occasional anomaly of this type might be seen as unimportant, need not occur. We wish to be able to say, even if in the course of a lifetime we undertake one sampling experiment and one only, or organize one lottery and one only, that in doing so we can act rationally.

We must re-emphasize that the assumption we make presupposes a three-fold system of apparatus, instructions, and operator, and that in consequence *fortuity* in the sense and circumstances in which we shall use the word will be regarded as *relative to the sensory equipment of the operator*. We shall regard, for example, the outcome of the tossing of a coin as fortuitous in an unqualified sense, only insofar as we believe that a task (the inducing of the result *heads*, say) is implicitly set which lies quite beyond the ability of *all* potential operators to fulfil. If an extraterrestrial visitor were to show himself able to spin a coin several feet in the air so that it invariably landed with head uppermost we should reasonably admire his ability. We should not be led to deprecate such laws of chance as we can now formulate.

By the above analysis we see that a variety of simple actions can provide the means for evolving stochastic sequences. We require only that the task of inducing a determinate outcome can be made exacting to any desired extent. We are released from the necessity for relying on traditional, and possibly inconvenient, devices. Further and more important, with an understanding of the rationale of any device chosen, we shall be able to see how its necessary inadequacies are remediable.

III. Probability Rules assuming Perfect Apparatus and Absolute Fortuity

A chance apparatus may prove deficient in two respects. In the first place, and trivially, physical difficulties may intervene before the boundary of the domain of fortuity is reached. For example, we can conceive of a coin too heavy to be used for our present purposes. In the second place we must recognize that there are no means by which the apparatus used can be made physically perfect. All dice are slightly asymmetrical, and all tops a little unbalanced. In discussing the fundamental rules of probability, however, we shall find it convenient to assume provisionally that the apparatus used is ideal.

With ideal apparatus we can in all cases, as in the example of the top, map the outcomes of individual trials on a line which is divided at equal intervals. When outcomes with a particular specification are distinguished a repeated pattern is imposed on this line. Thus Fig. 1 will depict a section of the domain of fortuity corresponding to the chance procedure in which the apparatus used is a symmetrical top divided into six equal sections, one of which is painted black, the outcome being scored *black* if the black sector comes to rest in contact with the surface on which the top is spun, the angle through which the top passes before it comes to rest being ϕ.

In this example, the mapping is linear. Extending the notions discussed, we can conceive of situations in which the mapping is on a plane, and in which a repeated pattern is imposed upon a meshwork of equal rectangular cells. In such situations we issue a two-fold instruction to the operator, who will have, in effect, to compound two stochastic actions. For example (Fig. 2) we may instruct the operator to spin a needle of length $2a$ across an area laid out in parallel strips each l units wide, imparting to it a minimum angular and a minimum linear momentum; or alternatively—and the effect will be the same—we can require the operator first to project the needle directly across the surface and then to spin it, pivoted at the point at which its centre has earlier come to rest. The cells of the meshwork on which outcomes are to be plotted will have sides of lengths l and π, and the repetitive pattern will be of the form exhibited in the diagram, θ being the angle through which the needle is spun and x the distance through which the centre of the needle travels before it comes to rest. This affords the celebrated problem of *Buffon's needle*.

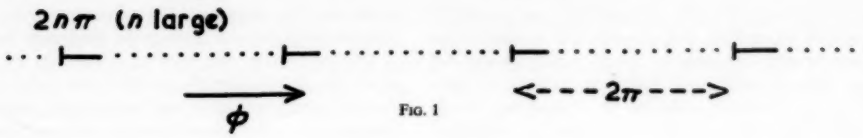

FIG. 1

FIG. 2

The meshwork of the mapping of any two-dimensional fortuitous behaviour must comprise identical rectangular cells. This follows because the primitive notion of fortuity we have prescribed requires that the picture in the domain of fortuity should be unaltered when the skeleton and pattern is shifted a whole cell-length along either axis. Accordingly we define the *probability* of a fortuitous event unambiguously, if we say (using the two-dimensional case in illustration) that it is the ratio of the area representing that event in a single cell of a rectangular mapping to the total area of the cell. Thus in Fig. 1, where the mapping is linear,

$$\text{Prob(event)} = \frac{1}{6};$$

whereas in Fig. 2,

$$\text{Prob(event)} = \frac{2}{\pi l} a \int \sin\theta \, d\theta = \frac{4a}{\pi l}.$$

This definition, in conjunction with the primitive axiom we have adopted, at once establishes the probability of an event which is certain as unity. It also yields directly the addition rule for compound events, the rule establishing probabilities in conditional sequences, and the product rule for successive trials. For example, the probability that a head results from adequate tossing of an ideal coin is $\frac{1}{2}$. In order to see that the probability of two heads occurring in successive tosses is $\frac{1}{4}$ we need to note that, since the individual outcome is fortuitous, we may alter the labelling of the coin *between* tosses without altering any essential feature of the procedure. Accordingly, the outcomes *HT*, *HH*, *TH*, and *TT* are of precisely equal standing, appear as equal squares in the mapping of the outcome of the double trial, and are each, therefore, to be taken to occur with probability $\frac{1}{4}$.

It must be emphasized that, in taking up this position, we advance beyond the classical notion of geometrical probability, completely avoiding its ambiguities in making our definition ultimately depend on a precise specification of the instructions issued to the operator. Equally, we avoid the obverse difficulty which results from the adoption of a frequency definition.* Furthermore, other

* See in particular the discussion of the needle problem by von Mises (von Mises, 1931).

approaches afford little foundation on which a discussion of errors due to the use of imperfect apparatus can be based. Such a discussion we must now proceed to. But our notion of probability as it stands subsists in a shadowy world of interphenomena. We must seek to establish some tangible consequence of the ideal theory we have envisaged. The only approach open to us here lies in the examination of the theoretical probabilities associated with compound events occurring within large numbers of trials. As a result of this examination we shall reasonably be led to see numerical values for probabilities reflected in observable *proportionate frequencies*. In no sense, however, shall we be tempted to endorse the inverse procedure of interpreting observed limiting frequencies as probabilities.

IV. Verifiable Consequences—Bernoulli's Theorem

Let us suppose that an experiment is repeated in a laboratory and that circumstances are so controlled that a particular result can be produced with unfailing success. Let us suppose further that, unknown to the experimenter, an assistant adopts the following procedure. At each repetition of the experiment he tosses a coin (which we may suppose to be ideal) ten times; if ten heads result he alters the conduct of the experiment so as to produce a contradictory result. If we can reasonably suppose that the possibility of deception thus introduced will have little effect on any honest appraisal of the results of the experiments—and this would surely be difficult to deny—we can reasonably accept, as effectively determinate, results in a *probabilistic* experiment which are predicted as occurring with probability very near to unity. One proviso only is needed; we are obliged faithfully to record the results of any empirical investigation of a chance mechanism, with no selection subsequent to the actual performance.

With this understanding we can derive verifiable and non-trivial consequences of the fortuity axiom. Consider the fortuitous event E with probability p produced by means of ideal apparatus. By virtue of the addition and multiplication rules, it follows from a known proposition concerning the terms of binomial series that, if n trials are performed, at each of which the event E may occur, then given δ and ε, it is possible to find a number N_o such that for all $n > N_o$.

$$\text{Prob}\left\{\left|\frac{m}{n} - p\right| > \delta\right\} < \varepsilon,$$

where m is the number of occurrences of E. This is commonly known as Bernoulli's theorem. It is only at this point that we need to invoke it. If we can agree on an acceptably small value for ε we can then say that, in a large number of trials, using ideal apparatus, the fortuitous event E, whose probability is p, will occur with a proportionate frequency approximately equal to p, and in a particular inquiry we can define numerically what we mean by the term *approximately* and what we mean by the term *large*.

V. Non-Ideal Systems

We proceed to examine the consequences of using imperfect apparatus for the elaboration of random sequences. If a biassed die is repeatedly thrown, we shall in general still be able to assert that sequences evolved are *random*, since it may still be quite beyond the operator's powers to induce determinate outcomes. But we shall no longer be able to associate the probability $\frac{1}{6}$ with the occurrence of any particular outcome. We must find out just what type of statement can be made in a situation such as this.

The simplest conceivable deviation from the ideal result occurs in situations where all predicted probabilities are deflected by fixed amounts from their true values. Thus the probability of throwing heads with a biassed coin might be supposed to become $\frac{1}{2} + \varepsilon$ instead of $\frac{1}{2}$. Have we, however, any right to make the assumption that a physical bias is in this way unambiguously representable by a numerical probability value? Clearly not, for to do so is merely to replace one ideal probabilistic system by another. In any particular situation, such as we envisage here, the reasons for rejecting the notion are especially cogent. The departures from theoretical perfection we are considering will be very small, since we shall be supposing that we have made every possible attempt to ensure the symmetry of the apparatus used. Such departures, which as we shall show later may well be important, we must assume will remain concealed after any attempt to verify their existence in trials of realizable length. In these circumstances we shall have no right to postulate anything less than the most general specification of departure from the ideal set-up.

This conclusion is suggested by a consideration of the behaviour of stochastic sequences the apparatus for the manufacture of which is knowingly biassed. Consider, for example, the case of coin-tossing. The coin will usually come to rest within a second or two of landing on a level surface. But occasionally, especially if the axis round which the coin is made to

spin whilst in the air is nearly vertical, the coin will spin topwise for a while after touching the surface. Normally this will not matter, but if we bias the coin (a penny) by soldering to the tail-side a sixpence, then, although the frequency of tails will be little affected in ordinary throws, in the sub-sequence of throws in which surface-spinning occurs the outcome will almost always be tails. If we impart a bias by attaching a disk rather lighter than a sixpence, we might suppose the situation to be represented roughly as follows:

In normal throws, Prob(tails) $= \frac{1}{2}$;

In throws resulting in surface-spinning, Prob(tails) $= \frac{1}{2} + \varepsilon$.

Now the operator will alter his method of throwing from time to time and eventually he may learn how to ensure that a throw with associated probability $(\frac{1}{2} + \varepsilon)$ results. Unless we alter our instructions to the operator, the most we can say of this system is that the probability of the result labelled *tails* alternates at perhaps unknown intervals between $\frac{1}{2}$ and $\frac{1}{2} + \varepsilon$. Within these limits the operator may be able consciously to impose on the system any binary pattern of underlying probabilities he wishes. Even if he does not possess the clue which would tell him how to vary the probability at will, we must still allow that such patterns may result. We have here an indication of what sort of departure from the ideal we should properly seek to account for.

The most general behaviour of this type we can envisage will occur when the probabilities concerned vary unpredictably from trial tc trial but remain always within certain specified limits. Let us suppose that in the system with which we are concerned each of r outcomes $A_1, A_2, \ldots, A_r$, should ideally result with probability $\frac{1}{r}$. Then, at the j^{th} trial of a sequence, we shall assume that:

$$\text{Prob}(A_i) = \frac{1}{r} + \frac{\delta_{ij}}{r}, \text{ where } \sum_i \delta_{ij} = 0$$

$$\text{and } \left|\delta_{ij}\right| < \delta, \text{ for all } i \text{ and } j.$$

We shall refer to δ as the *maximum proportional bias* of the system.

In application to the lottery problem, we must be able to assign an unequivocal value to the maximum proportional bias of the system which is specified by the physical apparatus, instructions, and operator employed. Such an assessment must be based on an examination of observed frequencies. However, this in no way contradicts our initial requirement that the primary grounds for supposing a sequence to be stochastic must be *a priori*. If at each assay the operator could perform an infinite number of trials, we should be able to assert that in no case could he succeed in educing limiting ratios greater than $\frac{1}{r} + \frac{\delta}{r}$ or less than $\frac{1}{r} - \frac{\delta}{r}$.

With a finite but large number of trials, at each assay we should be able with almost complete assurance to say, by reason of Bernoulli's theorem, that he will not succeed in educing limiting ratios outside a range

$$\left[\frac{1}{r} + \frac{\delta}{r} + \Delta_1, \frac{1}{r} - \frac{\delta}{r} - \Delta_2\right].$$

It follows that, if we can feel confident that the operator cannot educe in a fixed number of trials a proportionate frequency outside the range $\frac{1}{r} \pm \Delta$, we are justified in accepting the maximum proportional bias of the system as less than $r\Delta$.

In the case of coin-tossing sufficient is known to suggest that we are dealing with a truly fortuitous procedure. Assuming it implicitly specified that a minimum initial angular momentum should be imparted to the coin, the *a priori* requirements for the production of stochastic events are satisfied, and on the basis of experience we can reasonably say that in batches of 2,000 tosses no operator will succeed in obtaining a proportion of heads outside the range $\frac{1}{2} \pm \frac{1}{10}$.

If we can accept this, we have an adequate specification of a simple stochastic system whose maximum proportional bias is $\frac{1}{5}$.

We can thus resolve the simplest form of the lottery problem confident that the error we commit is less than 20 per cent. This does not immediately appear as a very striking achievement. Bút we shall proceed to show that, by using so crude and elementary a finding, we can resolve the lottery problem in its most general form.

VI. The Involution* of a Stochastic Sequence

In the ideal system, the operator cannot alter any long-run numerical property of a derivative sequence. But, if we admit the possibility of "*bias*", we allow him a limited opportunity for exercising a choice, or

* The simplest case of what we here call *involution* occurs when two coins are tossed, the labelling of the sides of one as *heads* or *tails* being switched if the other shows up, *e.g.* as heads. This procedure was suggested by Laplace. It has been pointed out to me that the procedure has been re-introduced recently, being referred to as *filtering* or *filtration*. It will be clear from what follows why the term *involution* is preferred. There exists no real analogy with harmonic analysis in the present context.

admit that he can in some way unthinkingly influence the course of events. It is natural to inquire whether we may use the element of fortuity necessarily present in the sequences evolved to vitiate such activity. In other words, we may anticipate that it is possible so to turn a sequence in upon itself as to rid it of irregularities imposed on it by the physical imperfections of the apparatus used.

Let us suppose, as above, that we have a stochastic system for which at every trial each of r events ideally occurs with probability $\frac{1}{r}$, and in which the maximum proportional bias is δ. We must allow that, within the limits afforded by the specification of δ, the operator has the knowledge to impose any underlying stochastic pattern he wishes upon the sequence which results from a set of ordered trials. We may divide these patterns into two classes according to whether we do or do not impose one condition on the circumstances under which the trials are carried out.

(*a*) We may suppose the trials to be undertaken with no restrictions imposed beyond those implicit in the specification of the apparatus used and the primitive instructions given to the operator.

(*b*) We may require that the operator should be kept in ignorance at each stage of the outcome of earlier trials, results being noted down by an independent recorder.

This distinction is sharp. We can conveniently refer to the two types of sequence respectively as *observed* and *blind*. In the case of blind sequences, any pattern imposed by the operator will have been so imposed independently of any knowledge he possesses of the outcome of any single trial within the sequence under consideration. Using the word in this special sense, we can, therefore, refer to trials in a blind sequence as *independent*.

Consider now the case $r = 2$. Let us suppose the two possible outcomes of the individual trial to be labelled 0 and 1. We can consider a sequence of trials to be condensed into a sequence of pairs of events. Each pair will be labelled either 0,0; 0,1; 1,0; 1,1. Let us first consider the case where the initial sequence is blind. We cannot ensure that for a particular pair the operator will not behave so that

at the first trial, $\text{Prob}(0) = \frac{1}{2} + \frac{\delta}{2}$

and at the second trial, $\text{Prob}(0) = \left(\frac{1}{2} - \frac{\delta}{2}\right)$.

In this case, by the product rule,

$$\text{Prob}(0,0) = \frac{1}{4} - \frac{\delta^2}{4};$$

$$\text{Prob}(0,1) = \frac{1}{4} + \frac{\delta}{2} + \frac{\delta^2}{4};$$

$$\text{Prob}(1,0) = \frac{1}{4} - \frac{\delta}{2} + \frac{\delta^2}{4};$$

$$\text{Prob}(1,1) = \frac{1}{4} - \frac{\delta^2}{4}.$$

Equally we cannot ensure that he will not behave so that, at each individual trial,

$$\text{Prob}(0) = \frac{1}{2} + \frac{\delta}{2},$$

so that:

$$\text{Prob}(0,0) = \frac{1}{4} + \frac{\delta}{2} + \frac{\delta^2}{4};$$

$$\text{Prob}(0,1) = \frac{1}{4} - \frac{\delta^2}{4};$$

$$\text{Prob}(1,0) = \frac{1}{4} - \frac{\delta^2}{4};$$

$$\text{Prob}(1,1) = \frac{1}{4} - \frac{\delta}{2} + \frac{\delta^2}{4}.$$

These are the two extreme cases. We can say therefore that the maximum proportional bias relating to paired events is approximately doubled.

Consider now the expression

$$\text{Prob}(0,0) + \text{Prob}(1,1)$$

in the blind sequence. If δ_1, and δ_2 are the effective biasses at the two loci in a representative pair ($|\delta_1|, |\delta_2| < \delta$),

then $\text{Prob}(0,0) + \text{Prob}(1,1)$

$$= \left(\frac{1}{2} + \frac{\delta_1}{2}\right)\left(\frac{1}{2} + \frac{\delta_2}{2}\right) + \left(\frac{1}{2} - \frac{\delta_1}{2}\right)\left(\frac{1}{2} - \frac{\delta_2}{2}\right)$$

$$= \frac{1}{2} + \frac{\delta_1 \delta_2}{2},$$

which is a maximum when $\delta_1 = \delta_2 = \delta$, and a minimum when $\delta_1 = -\delta_2 = \delta$.

Thus, $\text{Prob}(0,0) + \text{Prob}(1,1) = \frac{1}{2} \pm \frac{\delta^2}{2}$.

That is to say, if we relabel like pairs 0 and unlike pairs 1, then, if δ is the maximum proportional bias of the original sequence, we obtain a fresh stochastic sequence in which the ideal probabilities of the two possible events are $\frac{1}{2}$, and the maximum proportionate bias of which is reduced to δ^2. We may repeat this procedure using groups of 3, 4, . . . elements in order to produce stochastic sequences

with maximum proportional bias as small as we please. With the sequence divided into groups of n, the maximum proportional bias of the condensed sequence becomes δ^n.

We are now in a position to see why it is essential to distinguish between blind and observed sequences. For, if we suppose that at the first position in a pair the operator employs a proportionate bias δ,

$$\text{Prob}(0) = \frac{1}{2} + \frac{\delta}{2},$$

and that at the second position he employs a proportionate bias of $+\delta$ if the outcome of the preceding trial has been 0, and $-\delta$ if the outcome of the preceding trial has been 1, then,

$$\text{Prob}(0,0) = \frac{1}{4} + \frac{\delta}{2} + \frac{\delta^2}{4};$$

$$\text{Prob}(0,1) = \frac{1}{4} - \frac{\delta^2}{4};$$

$$\text{Prob}(1,0) = \frac{1}{4} - \frac{\delta}{2} + \frac{\delta^2}{4};$$

$$\text{Prob}(1,1) = \frac{1}{4} - \frac{\delta^2}{4}.$$

Whence

$$\text{Prob}(0,0) + \text{Prob}(1,1) = \frac{1}{2} + \frac{\delta}{2}.$$

It follows that the involuted sequence will have the same maximum proportional bias as the initial sequence.

Consider now the more general case. Let us suppose, using $\pm$ as indicated above, that

$$\text{Prob}(A_i) = \frac{1}{r} \pm \frac{\delta}{r};$$
$$i = 1, 2, \ldots, r.$$

Let us again consider a pair of trials, supposing that the operator has to deal with maximum proportionate biasses $(\varepsilon_1, \ldots, \varepsilon_r)$ at the first trial, and $(\eta_1, \ldots, \eta_r)$ at the second trial, where

$$\sum \varepsilon_i = \sum \eta_i = 0;$$
$$|\varepsilon_i| < \delta, |\eta_i| < \delta; i = 1, \ldots, r.$$

We then have:

$$\text{Prob}(A_i, A_j) = \left(\frac{1}{r} + \frac{\varepsilon_i}{r}\right)\left(\frac{1}{r} + \frac{\eta_j}{r}\right).$$

Let us sum this over r values of the pair (i, j) so that each value of i and each value of j is represented once and once only. The terms linear in ε and η will vanish so that:

$$\sum_{r \text{ values}} \text{Prob}(A_i, A_j) = \frac{1}{r} + \frac{1}{r^2} \sum_{r \text{ values}} \varepsilon_i \eta_j.$$

Now the selection of groups of pairs in this way may be represented by the laying out of a Latin square with rows and columns labelled according to the numbering of the first and second members of the pair. We select from all possible pairs those corresponding to a particular labelling of cells within the square.

The maximum value of $\sum_1^n x_l y_l$

$$\text{subject to } \sum_1^n x_l = \sum_1^n y_l = 0;$$
$$|x_l| \leqslant X, |y_l| \leqslant Y$$

is kXY, where $k = n$ if n is even, and $(n - 1)$ if n is odd.

It follows that, in the general symmetrical stochastic sequence with probability parameters $\frac{1}{r} \pm \frac{\delta}{r}$, involution by Latin square in pairs yields a stochastic sequence with probability parameters

$$\frac{1}{r} \pm \frac{\delta^2}{r} \text{ or } \frac{1}{r} \pm \frac{r-1}{r^2} \delta^2,$$

according as r is even or odd. Involuting in groups of s by means of an s-dimensional Latin structure, a condensed sequence results with a maximum proportional bias of at most δ^s.

The legitimacy of the involution procedure, insofar as it relates to the minute biasses which are our primary concern, depends on the theoretical argument advanced above. But when the bias is great its effectiveness can easily be observed in the behaviour of observable frequencies.

In a sequence of 3,000 spins of a badly balanced top, with two segments, one labelled X, the author observed 1,799 (59·9 per cent.) Xs.

In 3,000 double spins with involution, he observed 1,480 (49·3 per cent.) effective Xs.

In 3,000 triple spins with involution, he observed 1,496 (49·9 per cent.) effective Xs.

In 3,000 quadruple spins with 4-fold involution he observed 1,512 (50·4 per cent.) effective Xs.

VII. Resolution of the Lottery Problem in the General Case

Let us suppose that we have initially a stochastic sequence in which r events $(A_1, A_2, \ldots, A_r)$ ideally occur with equiprobability and in which the maximum proportionate bias is δ. We may conveniently

regard r as even. In using such a sequence to select n objects at random from amongst N we need to note:

(*a*) That the *reduced* sequence, derived from the original sequence by deleting the occurrences of $(\delta - s)$ of the fundamental events, comprises a stochastic sequence of s events, say $(A_1, \ldots, A_s)$ in which the maximum allowable proportionate bias is δ.

(*b*) That the *involuted* sequence, obtained by involuting the fundamental sequence in groups of k, comprises another sequence of r ideally equiprobable events with maximum allowable proportionate bias δ^k.

(*c*) that the *expanded* sequence, obtained by labelling separately distinguishable groups of k events, comprises a stochastic sequence with r^k ideally equiprobable events and maximum proportionate bias $(1 + \delta)^k - 1$.

Any sequence required in sampling may be derived from the fundamental sequence by appropriate reduction, involution and expansion. Reduction will be most conveniently carried out at the last stage. It is necessary to inquire in what order involution and expansion should be carried out.

Where two-fold involution is followed by two-fold expansion, the initial sequence being divided into fours, the resulting sequence comprises equiprobable events with probabilities:

$$\frac{1}{r^2} \pm \frac{2\delta^2 + \delta^4}{r^2}.$$

Where expansion is followed by involution, we have to consider in more detail the probabilities associated with four successive trials.

Suppose the effective biasses at these trials are ε_{ik}, where $i = 1, \ldots, r$; $k = 1, \ldots, 4$; and $\sum_i \varepsilon_{ik} = 0$ for all values of k.

Then

$$\text{Prob}(A_l, A_m, A_u, A_v) = \left(\frac{1}{r} + \varepsilon_{l1}\right)\left(\frac{1}{r} + \varepsilon_{m2}\right)\left(\frac{1}{r} + \varepsilon_{u3}\right)\left(\frac{1}{r} + \varepsilon_{v4}\right),$$

and the probability of any individual event in the sequence obtained in following expansion by involution is this expression summed over r^2 terms in which each pair (l,m) is represented once and once only, and each pair (u,v) once and once only, that is to say over cells with a particular labelling in a Latin square containing r^4 cells. Now in any Latin square rows and columns may be interchanged so that one set of similarly labelled cells lies along the principal diagonal of the square. If we re-label the biasses and call the event associated with the principal diagonal E, then

$$\text{Prob}(E) = \sum_{i,j} \text{Prob}(A_i, A_j, A_l, A_j) = \frac{1}{r^2} + \frac{1}{r^2}\sum_{i,j} \varepsilon_{i1}\varepsilon_{i2}\varepsilon_{j3}\varepsilon_{j4} + \sum_{i,j} \varepsilon_{i1}\varepsilon_{j2}\varepsilon_{i3}\varepsilon_{j4},$$

on expanding and using the relations between the ε_{ij}.

Now, if we put $\varepsilon_{ij} = (-)^j \frac{\delta}{r}$, we get:

$$\text{Prob}(E) = \frac{1}{r^2} + \frac{2\delta^2}{r^2} + \frac{\delta^4}{r^2}.$$

This indicates that we shall lose nothing *vis à vis* our present problem if we involute first and expand afterwards.

Consider now the problem of selecting one object at random from amongst N. Let us suppose the initial stochastic sequence we have to use is binary, with probability parameter $\frac{1}{2} \pm \frac{\delta}{2}$, and, for convenience, that the proportionate accuracy required in the final outcome is also δ. If k is the degree of expansion required and r is the degree of involution, then k is the smallest integer for which $2^k > N$, and r is the smallest integer for which $(1+\delta r)^k - 1 < \delta$. That is to say, r is the smallest integer greater than

$$\frac{\log[(\delta + 1)^{1/k} - 1]}{\log \delta}.$$

For example, for $N = 1{,}000$: $\delta = \frac{1}{10}$, $k = 10$, $r = 2$,
and for $N = 50{,}000{,}000$: $\delta = \frac{1}{10}$, $k = 26$, $r = 3$.

Finally, consider the general problem of selecting n objects at random from amongst N, using the same initial sequence and requiring of the result the same degree of accuracy.

We wish to ensure that each of the $\binom{N}{n}$ possible choices is made with probability $\frac{1}{\binom{N}{n}}(1 \pm \delta)$.

With N large, it would be impracticable to label each combination and employ the same method as for selecting a single object.

Accordingly we label the population $1, \ldots, N$ as before, and select by means of a sequence with probability parameter $\frac{1}{N} \pm \frac{\varepsilon}{N}$, successively reducing it as items are withdrawn (we are supposing all along

that sampling is without replacement). The ideal probability of getting a particular n-fold sample can be expressed in the form,

$$\left(\frac{n}{N}\right)\left(\frac{n-1}{N-1}\right)\cdots\left(\frac{1}{N-n+1}\right),$$

so as to correspond with the procedure employed.

Possibly associated with each term of this product is a term $(1+\varepsilon)$ corresponding to the largest bias allowed. Accordingly, ε must be chosen so that

$$(1+\varepsilon)^n - 1 < \delta\,.$$

For the present purpose we can take $\varepsilon = \dfrac{\delta}{n}$.

It follows, k being defined as before, that, for the degree of involution required in the initial sequence, we have:

r = smallest integer for which $(1+\delta^r)^k - 1 < \dfrac{\delta}{n}$.

For example, if we take $\delta = \dfrac{1}{10}$,

$N=1{,}000$,	$n=50$;	then $k=10$, $r=4$.
$N=1{,}000$,	$n=500$;	then $k=10$, $r=5$.
$N=50{,}000{,}000$,	$n=1{,}000$;	then $k=26$, $r=6$.
$N=50{,}000{,}000$,	$n=25{,}000{,}000$;	then $k=26$, $r=10$.

The values for r we obtain are therefore sufficiently small for us to take it that for practical purposes the lottery problem is satisfactorily resolved.

VIII. General Considerations

In conclusion we may briefly review the approach we have taken. It is customary to classify theories of probability which have any pretence to direct application as deriving either from a subjective or from a frequency notion of probability. Both of these we have implicitly repudiated. In spite of this, however, and, although we do not seek in any way to release ourselves from the discipline of numerical appraisal which a frequency theory imposes, the starting point of our inquiry has essentially been in the subjectivist camp. In this context, however, *subjective* is perhaps an unfortunate term. What the non-frequency schools have in common is that they all relate probability in some way to ignorance; and since ignorance is a quality commonly unequally distributed amongst individuals, theories result which are subjective in application.

What we have sought to do is to discover an *objective* form of ignorance and to relate the concept of chance to it accordingly. Keynes (1921) remarks:

> When we attribute a coincidence to objective chance, we mean not only that we do not actually know a law of connection, but, speaking roughly, that there is no law of connection to be known. And when we say that the occurrence of one alternative rather than another is due to chance, we mean not only that we know no principle by which to choose between the alternatives, but also that no such principle is knowable.

The individual's limited capacity for sensory discrimination merely affords us the only means possible of establishing a communicable meaning for the phrase "necessary ignorance" implied in the above quotation.

The position of frequency theories *vis à vis* our own is, however, of some interest. A frequency theory of probability (in other words a statistical theory) has the crippling disadvantage, as far as our present purpose is concerned, that it cannot provide any constructive clue to guide us in initially setting about producing a random sequence. The suggestion it will offer is that one should, by trial and error, seek to find such a sequence as it occurs in nature, testing sequences one by one for stability of proportionate frequencies in long series of trials. Such a procedure may well be regarded as sufficiently distasteful to warrant the rejection of any theory which proposes it. But suppose that we do provisionally accept it and that we succeed in obtaining a method of randomization which yields proportionate frequencies approximately in accord with theoretical requirements. We are then in the same position as we found ourselves at the end of Section V (above); namely, we can solve approximately the simplest form of the lottery problem, but no more. But since we shall not have admitted any *a priori* notions which would enable us to make the necessary distinction between observed and blind sequencies we cannot, as we did in Section VI, proceed to invoke the procedure of involution to advance matters further. A sequence established as random in the methodology of naive empiricism can, we must presume, be taken to have no qualities beyond those established by the direct scrutiny of routine enumeration. Moreover, we cannot obtain an alternative resolution of the difficulty by extrapolation. We might, working on the basis of empirical inquiry, find ourselves able to say that the "probability" of a coin landing heads is *approximately* $\frac{1}{2}$. But, clearly, we can have no right to infer from this that the probability, for example, of a run of ten heads conditional on the occurrence of *some* run of ten identical throws is *approximately* $\frac{1}{2}$ also. An empirical theory of this type can summarize, but it cannot, in any strict sense of the word, predict.

Finally, since in the context of the therapeutic trial the lottery problem is only of interest insofar as it must be resolved before we can discuss the inverse sampling problem, we shall mention briefly the

relation borne by our present findings to the second question propounded at the beginning of this communication. If we admit the necessary existence of an assessable experimental error in any sampling procedure, we must require of an inverse theory that it shall be able to take such error into account. In my suggested approach to inverse problems (Wrighton, 1953), we associate with the design of any statistical experiment an appropriately small probability value. In the simple situation of single-stage sampling, the maximum proportionate error of this probability value will be identical with the maximum proportionate bias of the sequence used for the selection of the sample after it has been adjusted by involution. It follows that, if the probability of drawing a particular sample is

$$\frac{1}{\binom{N}{n}}(1 \pm \delta)$$

and the theoretical uncertainty level is ε, the "true" uncertainty level of the experiment is

$$\varepsilon(1 \pm \delta).$$

Summary

(1) The object of this paper is to elucidate the notion of probability as associated with games of chance and with the procedure of random sampling as used in a therapeutic trial.

(2) We see the notion of probability as arising in situations where an operator performs a task for which his limited capacity for sensory discrimination precludes a determinate outcome.

(3) We note the significance of unavoidable imperfections in the physical apparatus used and put forward a method for rendering them ineffective.

(4) On the basis of the particular meaning we thus attach to the term *probability*, we advance a procedure for dividing N individuals at random into groups of n and $N-n$ members.

(5) The *probability* of any particular allocation will be assignable as

$$\frac{1}{\binom{N}{n}}(1 \pm \varepsilon),$$

where ε is small and prescribed and represents the maximum absolute error we are prepared to allow to the procedure.

(6) The full details of the proposed procedure are given in the text. Here it suffices to emphasize in what way the approach is fundamentally different from current recipes for random choice.

(7) Any fixed table of random numbers is the record of a unique sample ostensibly specified by a randomizing procedure; it cannot therefore be the basis for a continuing randomizing process.

(8) Methods hitherto proposed to effect such an end, though immune to this objection, fail to guarantee the accuracy of the probability values they lead to, and in many cases do not admit of complete specification.

(9) By circumscribing the notion of chance employed in this context, the procedure here discussed claims to be free from the foregoing defects.

References

Kendall, M.G. (1945). "The Advanced Theory of Statistics", 2nd ed., vol. I, p. 196. Griffin, London.
—— and Smith, B. Babington (1938). *J. roy. statist. Soc.* **101**, 147.
Keynes, J. M. (1921). "A Treatise on Probability". Macmillan, London.
Mises, R. von (1931). "Wahrscheinlichkeitsrechnung". Deuticke, Vienna.
Popper, K. (1935). "Logik der Forschung". Springer, Vienna.
Wrighton, R. F. (1953). *Acta genet. (Basel)*, **4**, 312.
—— (1955). *British Journal of Preventive and Social Medicine*, **9**, 135.

Brit. J. prev. soc. Med. (1957), **11**, 73–78

A NOTE ON THE PRACTICAL INTERPRETATION OF 2 × 2 TABLES

BY

J. H. EDWARDS

From the Department of Social Medicine, University of Birmingham

There have been intensive studies of the problem of whether the proportions into which any set is doubly dichotomized could reasonably be due to chance, and numerous tests and tables are available, although there is still disagreement about some fundamental assumptions underlying most test procedures. In practice, this problem should not commonly arise, since, outside genetics, there is almost always an *a priori* expectation of some interaction, and, in such cases, a significance test is merely a method of demonstrating whether the size of the sample is sufficient to exhibit this interaction. In most cases the experimenter or observer is mainly interested in the extent of the interaction, and its confidence limits.

There are three situations in which data are commonly assembled into 2 × 2 tables.

(1) The classical situation of paired attributes.

(2) Two populations, defined by some attribute, further dichotomized by the more or less arbitrary division of some continuously distributed variate.

(3) The dichotomizing of two continuously distributed variates.

Case No. 1

In the case of paired attributes, it is necessary to distinguish between two distinct situations, which may be termed simple sampling and compound sampling.

In simple sampling, a supposedly random sample is drawn from some population, and then classified into a 2 × 2 table by two pairs of exhaustive and exclusive attributes.

In this case, taking the classification

a	b
c	d

,

the most meaningful expression of any difference is that between the proportions

$$\frac{a}{a+b} \text{ and } \frac{c}{c+d} \text{ or } \frac{b}{a+b} \text{ and } \frac{d}{c+d}.$$

Any difference in proportion may be most meaningfully expressed as either a ratio or a difference depending on circumstances.

Consider the case when a difference is relevant (for example, the cure rate between two remedies).

If a, b, c, d represent

	NOT CURED	CURED
NOT TREATED	a	b
TREATED	c	d

,

the appropriate estimator is, of course,

$$\frac{d}{c+d} - \frac{b}{a+b},$$

with variance

$$\frac{ab}{(a+b)^3} + \frac{cd}{(c+d)^3}.$$

On the other hand, if the ratio is relevant (for example in assessing the value of an immunization procedure against a disease), when a, b, c, d refer to

	NOT ATTACKED	ATTACKED
NOT IMMUNIZED	a	b
IMMUNIZED	c	d

,

the appropriate estimator is

$$\frac{\dfrac{d}{c+d}}{\dfrac{b}{a+b}},$$

the sampling variance of which is not simple.

In practice, b and d are often so small compared with a and c that the approximation:

$$\frac{\frac{d}{c+d}}{\frac{b}{a+b}} \simeq \frac{d/c}{b/a} = \frac{ad}{bc}$$

will allow a simple solution with trivial bias.

In compound sampling, a subset of a population specified by some attribute, for example a disease, is classified in terms of another pair of exhaustive and exclusive attributes, and then the parent population is sampled and classified similarly. The situation considered by Woolf (1955) is that of blood groups and disease. When the subset represents an extreme minority (as is usual in disease), simple and compound sampling differ trivially.

In this case, the most meaningful expression of the interaction is the liability of one attribute to be related to another relative to the liability of the opposite of the first attribute: a relationship specified by the ratio of the products of the diagonals (the cross-ratio). For example, if one considers a random sample of p_t patients with some disease who smoke, and p_o who do not, compared with q_t and q_o controls respectively, randomly selected from the parent population, the relative liability of smokers to this disease compared with that of non-smokers is

$$\frac{p_t/p_o}{q_t/q_o} = \frac{p_t q_o}{q_t p_o}.$$

Woolf (1955) pointed out that the natural logarithm of this ratio is a maximum likelihood estimator with sampling variance

$$\frac{1}{p_t} + \frac{1}{q_t} + \frac{1}{p_o} + \frac{1}{q_o}.$$

Haldane (1956) has shown that an almost least biassed estimator is given by

$$\log_e \frac{(p_t + \tfrac{1}{2})(q_o + \tfrac{1}{2})}{(q_t + \tfrac{1}{2})(p_o + \tfrac{1}{2})},$$

with variance

$$\frac{1}{p_t + 1} + \frac{1}{q_t + 1} + \frac{1}{p_o + 1} + \frac{1}{q_o + 1}.$$

This provides an elegant and meaningful interpretation of such data, the confidence limits can be calculated directly, and also if desired the critical ratio on the null hypothesis. Being efficient, the unbiassed combination of data from various sources is practicable, and this does not depend on any equi-proportion in the expectation of the marginal totals: nor is the relevance of the subset of samples of identical marginal totals invoked.

The classical method of assessing significance is by the estimator

$$\frac{ad - bc}{N^2}$$

which has, on the null hypothesis, an expectation of zero and a sampling variance

$$\frac{m^4}{N^5},$$

where m is the geometric mean of the marginal totals. Although efficient in the limiting case of no interaction, this estimator has no intelligible meaning in most applications, and, in all three cases considered above, its use in assessing the significance of sets of data from various sources will lead to spurious heterogeneity unless the marginal totals are equi-proportional.

If these three methods of estimation are compared on one sample, and the significance of any discrepancy from the null hypothesis assessed, they will be found to give results which differ only trivially.

For example, consider the 2 × 2 table:

750	250
700	300

ESTIMATOR	VALUE	CRITICAL RATIO
$\frac{ad - bc}{N^2} \pm \left(\frac{m}{N}\right)^2 \frac{1}{\sqrt{N}}$	$+0{\cdot}01250 \pm 0{\cdot}004992$	2·504
$\log_e \frac{ad}{bc} \pm \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$	$+0{\cdot}2513 \pm 0{\cdot}1004$	2·504
$\frac{b}{a+b} - \frac{d}{c+d} \pm \sqrt{\frac{ab}{(a+b)^3} + \frac{cd}{(c+d)^3}}$	$-0{\cdot}0500 \pm 0{\cdot}01994$	2·508

CASE No. 2

When some variate, common to two populations defined by attributes, is dichotomized, the most meaningful expression of the difference is the difference in units of standard deviation between the means of the population (assuming that the standard deviation is the same in both populations). If the dichotomy is not into grossly unequal parts, a close approximation to this difference may be computed, with confidence limits, by making use of the approximation of the first derivative of the logistic function

$$\left(\frac{1}{2(1 + \text{Cosh } x)}\right)$$

to the normal curve.

The method is essentially the comparison of medians, a usually less efficient but simpler alternative to the comparison of means. Although relatively inefficient in the case in which the parent distribution is normal, it may be more efficient in leptokurtic distributions and has the particular advantage of being less influenced by either the inclusion or the rejection of outliers, a feature which may be of great advantage in biology, and particularly in the biology of man.

In the case of random sampling from two populations with a normally distributed variate, the variances being equal, but the means differing by t units of standard deviation, the 2 × 2 table may be considered in the following terms where the numerical values of a, b, c, d are proportional to the areas in the Figure.

	−	+
POPULATION 1	a	b
POPULATION 2	c	d

In Population 1, the dichotomizing line is distant say t_1 from the median where

$$\frac{a}{b} = \frac{\int_{-\infty}^{t_1} \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}} dt}{\int_{t_1}^{+\infty} \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}} dt}.$$

t_1 has the sampling variance of

$$\frac{ab}{(a+b)f_1^2}$$

where f_1 is the ordinate t_1 units from the median.

Now, to a close approximation,

$$t_1 = \frac{\sqrt{2\pi}}{4} \log_e \frac{a}{b},$$

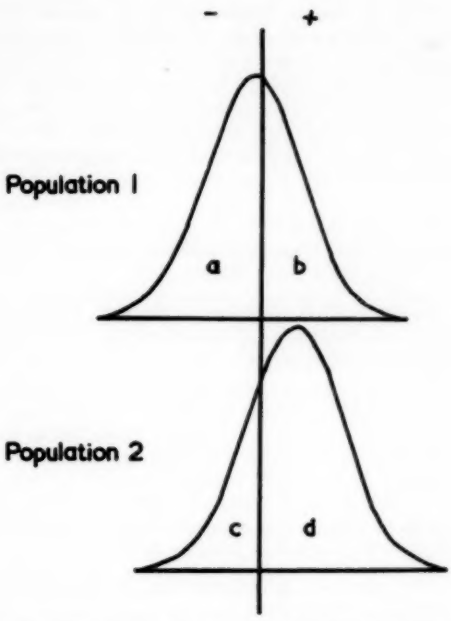

FIGURE.—Interpretation of a 2×2 table as an arbitrary dichotomy of two normal distributions.

$$\text{and} \quad f_1 = \frac{4}{\sqrt{2\pi}} \frac{ab}{(a+b)}.$$

So that the variance of the estimate is:

$$\frac{ab}{(a+b)\left(\frac{4}{\sqrt{2\pi}} \frac{ab}{(a+b)}\right)^2} = \frac{\pi}{8}\left(\frac{1}{a} + \frac{1}{b}\right).$$

Similarly,

$$t_2 = \sqrt{\frac{\pi}{8}} \log_e \frac{c}{d}, \text{ with variance } \frac{\pi}{8}\left(\frac{1}{c} + \frac{1}{d}\right).$$

So that the difference between the medians, $t = t_2 - t_1$,

$$= \sqrt{\frac{\pi}{8}}\left(\log_e \frac{c}{d} - \log_e \frac{a}{b}\right),$$

$$= \sqrt{\frac{\pi}{8}}\left(\log_e \frac{bc}{ad}\right),$$

with variance

$$\frac{\pi}{8}\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right).$$

This, of course, differs only in a constant term from Woolf 's method for estimating the interaction of attributes. It gives a direct answer in meaningful terms and, provided the disproportions are not too extreme (*e.g.* that the smallest tail is not dichotomized much more than one standard deviation from its mean), the bias is slight and the efficiency high enough to be useful.

The Efficiency Compared with the Estimates of the Mean

Given a single dichotomy, say into numbers a and b, the variance of the mean compared with that of the median is not less than:

$$\frac{\dfrac{1}{a+b}}{\dfrac{\pi}{8}\left(\dfrac{1}{a}+\dfrac{1}{b}\right)},$$

which reaches a maximum of $\frac{2}{\pi}$ (63·7 per cent.).

The efficiency is demonstrated in Appendix A, which also shows the bias.

The simplicity of the analysis, and of the collection and storage of the relevant data, will often outweigh the inefficiency and slight bias. For in experimental and observational data it will often be far easier to subdivide a continuous distribution into two than into many classes, especially when some "sieving" process can be employed, as in cutting a chromatogram, sorting eggs, diagnosing benign hypertension, exposing children to an 11+ examination, etc.; and where the amount of data collected can be several-fold increased by such measurement or classification a more precise estimate may be made of the difference between the medians than would be practicable for the difference between the means.

The method is particularly suited to the analysis of punch-card data, when the four sub-groups may be counted mechanically, and very large numbers handled. As in the previous case, a form of significance test in which the difference must be both appreciable and significant can easily be made by defining a "zone of triviality" which must be excluded from the confidence range.

Haldanes corrections* should be employed in estimating both the difference and its variance.

*In Haldane's opinion (personal communication) the term continuity correction is inappropriate in this context.

i.e.

$$\sqrt{\frac{\pi}{8}}\log_e\frac{(b+\tfrac{1}{2})(c+\tfrac{1}{2})}{(a+\tfrac{1}{2})(d+\tfrac{1}{2})},$$

and

$$\frac{\pi}{8}\left(\frac{1}{a+1}+\frac{1}{b+1}+\frac{1}{c+1}+\frac{1}{d+1}\right).$$

Case No. 3

In the third case, a 2 × 2 table may represent a bivariate population arbitrarily or otherwise dichotomized into quadrants: the most relevant question in relation to such data is the coefficient of correlation, and, because of the closeness of the normal curve to the first derivative of the logistic curve in any bivariate normal population, the cross-ratio is almost independent of the dichotomizing planes, providing they do not dichotomize into grossly unequal proportions, and this cross-ratio is therefore almost uniquely specified by the correlation coefficient. The natural logarithm of this cross-ratio is almost linearly related to Fisher's transformation of the correlation coefficient*.

$$\frac{\pi}{8}\log_e\frac{bc}{ad}\mathrel{\widehat{=}} z,$$

where

$$z=\tfrac{1}{2}\log_e\frac{1+r}{1-r},$$

and has a sampling variance

$$\left(\frac{\pi}{8}\right)^2\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}\right).$$

This estimate of z has a bias of the order $0{\cdot}02\,z^3$, which is of trivial consequence except when r is large, when the method is also unsuited through inefficiency.

The Efficiency of Estimates Compared with Complete Analysis

The efficiency, compared with a product-moment analysis of completely specified data is, in a large sample, not less than

$$\frac{\dfrac{1}{a+b+c+d}}{\left(\dfrac{\pi}{8}\right)^2\left(\dfrac{1}{a}+\dfrac{1}{b}+\dfrac{1}{c}+\dfrac{1}{d}\right)},$$

* A non-rigorous proof in a limiting condition is given in Appendix B.

or, in proportions when

$$a' = \frac{a}{a+b+c+d}, \quad b' = \frac{b}{a+b+c+d}, \quad \text{etc.},$$

$$\frac{64}{\pi^2\left(\frac{1}{a'}+\frac{1}{b'}+\frac{1}{c'}+\frac{1}{d'}\right)}.$$

The maximum efficiency (in the case $a = b = c = d$) is $\frac{4}{\pi^2}$ or 40·5 per cent. Although low, its simplicity gives it some advantages where vast data have to be extracted or coded with limited facilities or analysed with limited labour. Where data are only available in a 2 × 2 table, the method is far simpler than by the interpolation of tetra-choric functions and not less efficient.

The bias under various conditions, and also the cross-ratio when the dichotomies are through the means are shown in Appendix B. The calculations are based on Karl Pearson's tables for the volumes of quadrants of the bivariate normal surface.

As in the previous case, it is simple to test for "appreciable and significant" correlation by defining a zone of triviality. The cross-ratios, and their differences, tabulated in Appendix B, provide a rough method of estimating the correlation coefficient from the cross-ratio. Haldane's corrections should be used.

Summary

The conventional method of testing for interaction in a 2 × 2 table utilizes an estimator which has no direct meaning, so that it is not possible to dissociate the distinct questions whether an interaction is important and whether it is significant; nor is it usually possible to combine data from different sources without loss of information and bias.

In observational data, it is not usually reasonable to ask the question "is interaction present" (obvious exceptions occur with genetic linkage, blood groups and disease, and differences in response to *equally reasonable* remedies). In practice, interaction can usually be assumed, and its direction is usually obvious, and the only way in which statistics can assist in extracting meaningful information is by a meaningful estimator to which confidence limits may be calculated.

Some common situations in which data are acquired, or extracted, in a 2 × 2 form are considered, and some meaningful estimators, and methods of deriving their confidence limits, are discussed. One estimator provides a simple approximate method of estimating the regression coefficient from a 2 × 2 table pertaining to a normal bivariate population.

The concept of a "zone of triviality", as an aid to the detection of differences which are both significant and important, is introduced.

References

Haldane, J. B. S. (1956). *Ann. hum. Genet.*, **20**, 309.
Sheppard, W. F. (1897). *Proc. roy. Soc.*, **62**, 170.
Woolf, B. (1955). *Ann. hum. Genet.*, **19**, 251.

Appendix A

The bias and efficiency of calculating the difference between medians from the cross-ratio in a 2 × 2 table, one entry of which refers to a dichotomized normally distributed variate, are given below.

"δ" refers to the difference between the medians (or means) of the parent populations (in units of standard deviation), "d" the estimated difference, and "η" the efficiency. Two sets of calculations are made, one in the most efficient dichotomy midway between the means, (called the antimodal), the other through one mean (called the modal).

δ	Antimodal		Modal	
	d	η per cent.	d	η per cent.
0·2	0·2001	63·3	0·2004	62·8
0·4	0·4007	62·1	0·4029	60·4
0·6	0·6024	60·1	0·6098	56·4
0·8	0·8058	57·5	0·8235	51·0
1·0	1·0114	54·3	1·0454	44·3

Appendix B

The approximation

$$z \doteq \frac{\pi}{8} \log_e \frac{bc}{ad}$$

may be demonstrated as an equality in the limiting case of a bivariate normal surface of correlation r (when $r \to 0$) dichotomized through the means.

Consider the relative volumes:

a	b
c	d

$(a = d,\ b = c)$.

Then $r = \text{Cosine}\left(\frac{a}{a+b}\right)\pi$ (Sheppard, 1897).

$$= \text{Cosine}\left(\frac{1}{\frac{b}{a}+1}\right)\pi$$

Let $\frac{bc}{ad} = \left(\frac{b}{a}\right)^2 = w$.

When $r \to 0$, $w \to 1$, $1 - \sqrt{w} \to 0$, $z \to r$,

and neglecting second order terms

$$= \text{Cosine}\left(\frac{1}{2 - (1 - \sqrt{w})}\right)\pi$$

$$= \text{Cosine}\left(\tfrac{1}{2} + \tfrac{1}{4}(1 - \sqrt{w})\right)\pi$$

$$= -\text{Sine}\ \tfrac{1}{4}(1 - \sqrt{w})\,\pi$$

$$= -\frac{\pi}{4}(1 - \sqrt{w})$$

$$= \frac{\pi}{4}\log_e \sqrt{w}$$

$$= \frac{\pi}{8}\log_e w$$

$$= \frac{\pi}{8}\log_e \frac{bc}{ad}.$$

The cross-ratio of the evenly dichotomized bivariate normal surface, and estimates of r, based on the method described above, of a bivariate normal surface, of correlation r dichotomized (0, 0), (0, 0·5) and (0·5, 0·5) units of standard deviation from the means, are tabulated below.

r	Cross-Ratio (0, 0)	Estimates of r		
		(0, 0)	(0·5, 0·0)	(0·5, 0·5)
0·0	1·00 (0·29)	0·000	0·000	0·000
0·1	1·29 (0·38)	0·100	0·103	0·106
0·2	1·67 (0·52)	0·199	0·206	0·210
0·3	2·19 (0·73)	0·299	0·310	0·313
0·4	2·92 (1·08)	0·398	0·412	0·414
0·5	4·00 (1·70)	0·496	0·514	0·513
0·6	5·70 (3·00)	0·594	0·616	0·610
0·7	8·70 (6·37)	0·691	0·718	0·705
0·8	15·07 (20·51)	0·788	0·820	0·799
0·9	35·58	0·886	0·924	0·893

Brit. J. prev. soc. Med. (1957), **11**, 79–89

A CRITICAL EXAMINATION OF THE REPUTED PRIMARY INFLUENCE OF ABO PHENOTYPE ON FERTILITY AND SEX RATIO

BY

J. H. EDWARDS

From the Department of Social Medicine, University of Birmingham

Within the last decade, statistical evidence has accumulated that the ABO blood groups have selective influences other than those due to mismatched or unavailable transfusions, and an increasing literature has convincingly demonstrated a primary influence on at least one group of diseases in adults. Publications on the influence of the ABO phenotypes on fertility and sex ratio have made similar claims of apparent selective influences inexplicable by chance, and such claims have been widely quoted (*British Medical Journal*, 1954), and have been promoted to the attention of geneticists in the Bateson lecture on morphism (Huxley, 1955).

Some Difficulties of Statistical Analysis

The problem of analysis in the light of the exclusive and exhaustive classes of prior hypotheses of heterosis and foetal incompatibility would present considerable inferential problems even with a diallelic system, a Poissonian distribution of family size, random sampling, and a well miscegenated population. In practice, most data refer to triallelic or tetra-allelic systems; the distribution of family sizes is far from Poissonian; and random sampling has rarely been achieved. Populations with facilities for blood grouping are mostly cosmopolitan, with appreciable proportions of unusual blood group distribution (for example, Jews, Irish, or Negroes), and such minorities usually have different views, abilities, or opportunities in relation to their standard of living and family size.

The heterogeneity due to these factors is such that, in conjunction with statistical methods which ignore it, or which are assumed appropriate on the grounds that they have been helpful when applied to experimental data or to observational data consisting of far fewer observations, almost every set of data published can be analysed to show evidence of "significant" differences in fertility. Such differences would be expected to arise from the heterogeneity of the data itself, and the separation of the side-effects of a gene from the effects of other genes with which it is associated cannot usually be made on information implicit within any single set of data. Significance tests may be necessary, but they are certainly not sufficient, as evidence of any such effect.

While this multiplicity of claims is such as usually initiates the healthy development of a new subject, and while the task of the critic is comparatively easy, in that he has the advantage of comparing any two papers at least one of whose authors could not have read the other, it would appear to be worth while to try to analyse some of the inherent difficulties and to attempt to advance constructive suggestions.

Firstly, in the absence of any well-grounded prior hypotheses, other than the vague and contradictory expectations of foetal incompatibility and heterozygote advantage, interpretation must largely be based on posterior hypotheses. Consequently, tests of significance must be based on the whole data, and this tends to shield any heterogeneity limited to a few of many class frequencies. As it is quite possible that there may be systematic discrepancies limited to only one or two of many class frequencies, as would be expected in the various Rhesus parent-child combinations, it would seem valid to apply also a significance test to those classes most discrepant from the mean. As in a many-classed contingency table the ratio of the square of a discrepancy to its expectation (d^2/e) has, on the null hypothesis, the probability of a random variate from a normal population of zero mean and unit variance, significance tests may easily be applied using the probability distribution of the extreme deviates; these have been tabulated (May, 1952). Although only approximate, such methods would seem preferable to the highly fallacious methods of testing those classes with the most extreme difference for independence. There would, of course,

be a probability of almost pn of wrongly rejecting a null hypothesis with n independent tests at the pth level.

Secondly, conventional significance tests seem unsuited to such data. Methods dependent on the sampling distribution of attributes, as χ^2, may be useful when small numbers of observations are being analysed and fairly gross distortions of proportionality are necessary for the rejection of the null hypothesis, but are not suited to observational material in man where many thousands of cases are being compared. It is doubtful whether any null hypothesis of association of attributes tested on observational data derived from, say, 10,000 men would survive rejection at anything approaching $1-p$ times at the pth level of significance.

Most significance tests, including those dependent on the probability distribution of χ^2, are comparisons of the ratio of the square of a discrepancy to its variance, so that, as the size of the sample increases, any bias increases this ratio in proportion to the sample size. This leads to the paradox that the larger the series, the greater the caution necessary in accepting any apparent rejection of the null hypothesis. Consider a small systematic error γ. If the numbers are large, the discrepancy in a sample of number n will be $n\gamma$, with variance approximately n. Conventionally, it will be regarded as significant if $n\gamma$ exceeds $2\sqrt{n}$.

When $\gamma > \frac{2}{\sqrt{n}}$, this will occur "usually". When $\gamma = \frac{1}{\sqrt{n}}$, this will occur with a probability of about 16 per cent., and, when $\gamma = \frac{1}{2\sqrt{n}}$, of about 7·3 per cent.

For example, in a series of 10,000 cases, a 2 per cent. systematic error would "usually" appear significant at the 5 per cent. level, and in 60,000 cases at the one in a million level.

This is, of course, not a criticism of such methods of analysis as are designed to detect such discrepancies, but merely a demonstration of the confidence of freedom from small systematic influences which must be assumed before using such tests. The belief that the difficulty can be overcome by using higher levels of significance is not only fallacious, but harmful, in that very highly significant discrepancies are less readily dismissed as statistical artefacts.

In view of the obvious racial, social, and religious sub-grouping in man, it is doubtful whether such methods are applicable to his reproduction. Apart from errors of paternity and blood-grouping, misinterpretation of data may readily arise from such correlation nets as: Irish ancestry, high proportion group O, Catholicism, absence of contraception, larger families: Jewish faith, high proportion group B, atypical morbidity experience, absence of contraception, Niddah, highly assortative mating; Negro facies, high proportion group B, large family size, high neo-natal death rate.

And this misinterpretation will often appear to be confirmed by data from other regions. Further, race, religion, marital status, and parity may appreciably influence liability to become a blood donor.

A simple solution to some of the problems of a significance test which would be insensitive to small systematic bias would appear to be Woolf's efficient test of proportionality in a 2×2 table (Woolf, 1955), provided the null hypothesis is only regarded as rejected when the fiducial limits of this estimate exclude, not unity, but some arbitrarily determined range, say 0·95—1·05, which is considered unlikely to be exceeded by the bias which cannot be excluded. An essential, and unique, feature of this method is the possibility of combining data from diverse populations, without introducing bias or losing information. Haldane (1956) has published a refinement of Woolf's method which is very simple and reduces any bias due to the asymetrical sampling errors of small numbers. Data should therefore be subdivided as often as practicable: religion may be suggested as a simple attribute which is usually recorded on hospital notes (Penrose, 1955). This analysis will not detect very small discrepancies which are, by their nature, probably indiscernible in man. The impossibility of detecting small differences, while possibly discouraging the further accumulation of data on the ABO phenotypes, in which fertility is assessed by enumeration of the children of the fertile, should not discourage the collection of other data. Appreciable differences in the relative liability of different phenotypes may well occur in certain particular forms of fertility impairment or in sterility, and yet have escaped recognition. In these cases, Woolf's method of analysis would be particularly easy to apply.

As there are $\frac{1}{2}\,n(n-1)$ internal comparisons which can be made in an $n \times 2$ table (*i.e.* two populations classified into n phenotypes), the arbitrary adoption of some standard which would reduce the number of comparisons to $n-1$ would seem desirable. The recessive phenotypes (such as O, d/d) would seem in general most suitable, being fairly frequent in most races and of unique genotype. This solution is only applicable to problems which can be presented as an $n \times 2$ table, for example in relation to disease, to the sex ratio, and to sterility. It is not applicable to investigations involving

gradations in fertility. In such cases Student's "*t*" test is frequently used, but in view of the very poor approximation of family size distribution to normality or even to a Poissonian distribution, it would seem most unsuitable. The difficulty may be overcome with some loss of information, by dividing families into those above and below their median size, or some other percentile. More refined tests would require very exacting definitions of fertility, if equally reasonable tests were not to give opposing results.

Any absence of acceptable evidence is, of course, not negative evidence. A difference in fertility too slight to demonstrate (for example 1 per cent.) would probably be sufficient to have maintained this genetic variation throughout or beyond the time of man. Even the fairly intense selection against some heterozygotes of the Rhesus phenotypes would probably not be evident from statistical studies in the absence of a hypothesis derived from anecdotal evidence. Indeed, paradoxical discrepancies have been described in statistical studies relevant to such hypotheses (Glass, 1950).

A trend evident in many of these studies is that, while they have been explicitly made as a result of the theoretical necessity of differences with respect to genotype fitness (though not necessarily in phenotype fitness) to maintain such variation, the hypotheses advanced to explain such differences as have been claimed have usually been such as would lead to instability if they were real.

A Critical Examination of Published Work claiming an Association between Blood Group and Fertility.

Seven widely-quoted publications claim to have demonstrated an association between blood group and fertility.

(1) By assembling data from the literature, Waterhouse and Hogben (1947) inferred an intense selection against Group A foetuses of Group O mothers. This conclusion was criticized by Bennett and Brandt (1954), who considered that the differences in birth rank of Group A and Group O children of A/O marriages (Group A husband and Group O wife) and the differences in the proportions of Group O and Group A children of A/O and O/A marriages were "not other than trivial". Their rejection of the primary hypothesis of progressive maternal inhospitality to Group A foetuses by Group O mothers is on grounds of undisturbed birth order, and not, as is sometimes stated, of heterogeneity within the data.

Although the tests used by Bennett and Brandt are technically efficient, the absence of any significant discrepancy does not necessarily imply that the data do not contain evidence of selection through foetal incompatibility. Firstly, Haldane and Smith's efficient test of birth order disturbance on which Bennett and Brandt base their claim is a test for consistent trend. Consequently, it is not efficient as a test of trend in the birth order disturbances of later-born children if first-born are included and if they are (as is suggested by the histograms published by Waterhouse and Hogben) particularly prone to bastardy. If bastardy is far commoner in the first-born, the custom, once practised in the Highlands of Scotland (see Boswell's *Journal of a Tour of the Hebrides with Samuel Johnson*, 1786), of excluding the first-born from any ancestral status might be followed with advantage. Secondly, in relation to their claim of no significant disturbance in segregation ratio, it cannot be assumed that segregation recorded among the children of a mother of recessive group implies that the husband is heterozygous. The presence of a few large families, in which the husband was homozygous for the dominant gene, and in which one child was a bastard of recessive group, would greatly bias any data on segregation.

As an example of bias due to mistaken paternity, consider a large number of families, each consisting of n children of Group O mothers whose husbands are Group A, and let a small proportion (α) of children be not fathered by their mother's husband. If all families which contain a member of Group B or AB are excluded, one may consider the gene conferring the B antigen to be excluded from the universe under discussion.

Let the gene frequencies be p and r. The proportion of A/O marriages in which the husband is homozygous is $\frac{p^2}{1-r^2}$, and in almost αrn families there will be apparent segregation.

There will be $\alpha rn^2 \frac{p^2}{1-r^2}$ children of families containing a bastard; of these children $\alpha rn \frac{p^2}{1-r^2}$ will be of Group O and $\alpha rn\ (n-1) \frac{p^2}{1-r^2}$ of Group A.

If we ignore the relatively small bias from bastardy in families in which the husband is heterozygous, the segregation ratio (ratio of Group A to Group O

children in the families which segregate) will be:

$$\frac{\frac{npr(1-\alpha)}{1-r^2}+\frac{\alpha np^2r(n-1)}{1-r^2}}{\frac{npr(1-\alpha)}{1-r^2}+\frac{\alpha np^2r}{1-r^2}}=\frac{(1-\alpha)+\alpha p(n-1)}{(1-\alpha)+\alpha p}$$

If, as is probably common in European data, $p = 0{\cdot}3$, $n = 4$, $\alpha = 0{\cdot}05$, this ratio would be $1{\cdot}03$. In O/A marriages, the bias from bastardy would be far less, as spurious segregation could not occur.

Serious bias may also arise from grouping errors, particularly as they are usually systematic, and are commoner in children, particularly in young children (children of higher birth rank). Consider the simple universe defined above, ignoring bastardy. Assume the parents accurately grouped, and let the proportion of children of Group A misgrouped as O be β.

Then, of the A/O or O/A matings, there will be $\frac{p^2}{1-r^2}$ families of homozygous parents, of which almost $\frac{\beta np^2}{1-r^2}$ will be regarded as segregating, so that the segregation ratio will be:

$$\frac{\frac{npr(1-\beta)}{1-r^2}+\frac{\beta np^2(n-1)}{1-r^2}}{\frac{npr(1-\beta)}{1-r^2}+\frac{\beta np^2}{1-r^2}}=\frac{r(1-\beta)+\beta p(n-1)}{r(1-\beta)+\beta p}$$

If, for example $r = 0{\cdot}7$, $p = 0{\cdot}3$, $n = 4$, $\beta = 0{\cdot}05$, this ratio will be $1{\cdot}04$.

This apparent segregation, due to either bastardy, or grouping errors, or both, could clearly lead to considerable bias in interpretation.

(2) Allan (1953) reviewed the data of Waterhouse and Hogben and, concluding "that some other difference must be sought for the curious differences in fertility found in the sample", proposed an ingenious hypothesis of "inverse fertility", the fertility of each phenotype being supposed to differ from the mean by an equal and opposite amount in the two sexes. Before considering the evidence for this hypothesis, it is of interest to consider whether there is any justification for regarding these differences as curious.

One approach to this problem, which has sometimes been employed, is to test the combined null hypotheses of random mating and equal fertility by a test of unreasonableness of proportions or a test of independence relative to the marginal totals, assuming $(n-1)^2$ degrees of freedom where there are n phenotypes. This, however, is fallacious; in the presence of random mating, and of ascertainment independent of family size, the modal value of $\sum\frac{d^2}{e}$, where e is the expected number and d that number less the observed number, is not $(n-1)^2$ but $f(n-1)^2$, where f is the mean number of children per marriage. If $f = 3{\cdot}3$ and $n = 4$, this modal value will exceed the very highly significant value of $\chi^2_{(9)}$.

When information on relative fertility is provided only by those children who are in excess of a minimum number necessary for inclusion in the data, that minimum number, usually one, should be subtracted from all family sizes. A possible shortcoming of such data is that, if apparent sterility, that is childlessness at the time of collection of the data, differs in degree, rather than in nature, from relative infertility, series from which childless marriages are excluded may reveal apparent trends in fertility diametrically opposed to any which actually exist.

If the distribution of family size were Poissonian, the significance of any discrepancy from the expectation of uniform fertility could easily be tested, since

$$\sum_{1\ldots n}\frac{d^2}{e}$$

would be distributed as $\chi^2_{(n-2)}$.

Waterhouse and Hogben have published the family size of all O/A and A/O marriages, of which there were 453 with 1,068 children excluding the first, giving a mean of $2{\cdot}36$ and variance $8{\cdot}90$. It may reasonably be supposed, therefore, that in the data as a whole, the variance considerably exceeds the mean. Consequently, on the null hypothesis, the probability of obtaining a value of

$$\sum_{1\ldots n}\frac{d^2}{e}$$

or higher will not be less than the probability of exceeding $\chi^2_{(n-2)}$.

Even if "significant differences" could be demonstrated, these data would not necessarily imply any primary effect of blood group on fertility. In fact, even after excluding one child from each family, the differences in fertility (Tables I and II, opposite) are not beyond the reasonable expectations of chance and hardly form sufficient evidence for rejection of the null hypothesis.

TABLE I

NUMBER OF MARRIAGES AND CHILDREN DUE TO THEM IN EACH CLASS RELATIVE TO THE ABO SYSTEM

Blood Group of Husband		Blood Group of Wife				Total
		O	A	B	AB	
O	Marriages	225	244	54	19	542
	Children	779	835	171	56	1,841
A	Marriages	**209**	189	**63**	28	489
	Children	**686**	599	**196**	96	1 577
B	Marriages	**66**	**59**	17	8	150
	Children	**229**	**210**	62	30	531
AB	Marriages	**23**	**20**	**6**	9	58
	Children	**72**	**80**	**14**	24	190
Total	Marriages	523	512	140	64	1,239
	Children	1,766	1,724	443	206	4,139

Heterospecific marriages are shown in bold type.

TABLE II

VALUES OF d^2/e FOR FAMILY SIZE (LESS ONE CHILD) ACCORDING TO PARENTAL BLOOD GROUP

e represents the number of children expected on the null hypothesis and d the difference between observed and expected numbers

(From data assembled by Waterhouse and Hogben, 1947)

Blood Group of Husband	Blood Group of Wife				All Groups	
	O	A	B	AB		
O	1·423	0·693	0·698	1·255	0·729	6·17
A	0·304	2·369	1·418	0·092	2·795	
B	0·468	1·205	0·682	0·575	2·545	
AB	0·433	3·717	2·598	1·749	0·104	
All Groups	0·291	0·155	1·859	0·406	19·68	
	2·71					

When $\chi^2_{(14)} = 19{\cdot}68$, $0{\cdot}1 < p < 0{\cdot}2$
$\chi^2_{(3)} = 6{\cdot}17$, $0{\cdot}02 < p < 0{\cdot}05$
$\chi^2_{(3)} = 2{\cdot}71$, $0{\cdot}2 < p < 0{\cdot}3$

It would seem, therefore, that Allan's hypothesis receives little support from data in which the body of the table shows no significant differences in fertility and in which the marginal totals do not both differ significantly from expectation.

Allan has suggested that such inversions of fertility with sex may be a sufficient reason for the establishment and persistence of such variation, and it is of interest to consider this sufficiency on a simple model.

Consider such genotypes A_{ij} as are determined by autosomal alleles $A_1, A_2, \ldots\ldots A_n$, having frequencies $x_1, x_2, \ldots\ldots x_n$, and let the genotypes ${}_mA_{ij}$ and ${}_fA_{ij}$, where m and f imply male and female, have fertilities $1 + {}_ml_{ij}$ and $1 + {}_fl_{ij}$, and let the combined fertility of a mating pair be the sum of the independent fertilities.

Allan's hypothesis appears to imply:

$$(1 + {}_ml_{ij}) = (1 - {}_fl_{ij}).$$

Matings between genotypes which are alike will have relative fertility:

$$\begin{aligned}&(1 + {}_ml_{ij}) + (1 + {}_fl_{ij})\\ &= (1 - {}_fl_{ij}) + (1 + {}_fl_{ij})\\ &= 2.\end{aligned}$$

Matings between unlike genotypes will be: ${}_mA_{gh}$ and ${}_fA_{ij}$ or ${}_mA_{ij}$ and ${}_fA_{gh}$.

It is assumed that these will have equal frequency, a condition not necessarily implying either randomness of mating or unity of sex ratio. The former will have relative fertility: $(1 + {}_ml_{gh}) + (1 + {}_fl_{ij})$, and the latter: $(1 + {}_ml_{ij}) + (1 + {}_fl_{gh})$.

On Allan's hypothesis, ${}_ml_{gh} + {}_fl_{gh} = 0$, and ${}_ml_{ij} + {}_fl_{ij} = 0$. Therefore the mean relative fertility of these matings will be 4/2 or 2. That is, the numbers of each genotype will be constant in the next generation, whatever the gene frequencies (the equations being independent of $x_1, x_2 \ldots x_n$). The stability is therefore neutral. Allan's hypothesis is not, in this interpretation, a sufficient reason for a stable variation.

(3) Bryce, Jakobowicz, McArthur, and Penrose (1950), in an analysis of 7,856 consecutive mother-child pairs of known blood group in Australia, provided evidence of a relative deficiency of Group AB infants and irregularities in the distribution of blood groups among the children of Group A and B mothers. These observations were considered to be suggestive of differential fertility or foetal viability, or both concordantly.

The data were analysed by calculating the expected numbers in each mother-infant phenotype pair, on the assumption that the parents were a random sample, randomly mated, from a universe whose frequencies could be accurately inferred. For the body of their table

$$\sum \frac{d^2}{e}$$

is 24·26: this has been derived from fourteen class frequencies, only one parameter, the total number, being derived from the data, the other two parameters, specifying the gene frequencies, being derived partly from the data but mainly from other data assumed to have been assembled from the same universe. The heterogeneity in the body of the table is significant at thirteen degrees of freedom ($0{\cdot}01 < p < 0{\cdot}05$). However, differential fertility or foetal viability cannot definitely be inferred as a necessary cause, since there is evidence of a considerable, though not significant, discrepancy between the

expected and observed numbers of children of each blood group ($\chi^2_{(3)} = 6{\cdot}37$; $0{\cdot}05 < p < 0{\cdot}1$), the discrepancy being almost entirely due to a shortage of infants of Group AB, or an excess of Group AB mothers. The proportion of mothers of Group AB does not differ appreciably from that expected from the gene frequencies calculated from the other three phenotypes, nor from the proportion in the control population. The authors recognized the possibility of some systematic grouping errors, some AB infants being grouped as B, and comparison of the number of infants grouped as AB with the number expected from the gene frequencies calculated from the other phenotypes suggests that such errors may be appreciable. In view of the known difficulty of detecting the A_2 antigen in group A_2B infants, it appears reasonable to consider grouping errors* as a likely contributory cause of the observed discrepancies. The maximum likelihood estimate of this possible error, on the assumption that the children have the gene frequencies of the adult sample, is $0{\cdot}116 \pm 0{\cdot}045$ ($11{\cdot}6$ per cent. $\pm$ $4{\cdot}5$ per cent. AB infants grouped as B). If this error were entirely due to failure to detect the A_2 antigen, it suggests that about half the infants of group A_2B were grouped as B. If this estimate is used to calculate the expectations, the heterogeneity in the body of the table is considerably reduced,

$$\sum \frac{d^2}{e} \text{ being } 18{\cdot}7,$$

which, if distributed as $\chi^2_{(13)}$ (a second parameter having been estimated from the data), is not conventionally significant ($0{\cdot}05 < p < 0{\cdot}1$).

Although these data are very suggestive of a high fertility of Group AB women, a trend which could be a powerful influence in maintaining variation, the discrepancies are largely explicable on grounds of the technical difficulty of detecting in infancy the A_2 antigen in the presence of the B antigen. A follow-up of a series of infants presumed to be Group B would be of interest.

(4) Kirk, Kirk, and Stenhouse (1953) analysed the records assembled over a period of 4 years in a large Australian maternity hospital and claimed to have shown a highly significant trend in fertility, Group A women being more fertile than Group O between the ages of 25 and 35, and Group O women being more fertile before and after this period. They made no claim of any appreciable difference in overall fertility.

Their method consisted of fitting orthogonal polynomials of unstated degree to the unweighted difference between the mean number of pregnancies of Group A and of Group O women by each age. The regression line shown in their graph does not appear to have a high significance, apparently accounting for only about half the variance. Over half the children ascribed to women over 35 were from families exceeding six children, most of whom were presumably born before their mothers reached that age, so that the inference about the differential fertilities after age of 35 were made on data largely relevant to their fertilities before this age. In fact, the 271 women delivered between the ages of 35 and 40 are credited with 1,453 children and it is these children who form the major part of the data used to establish the difference in fertility in this age group. The problem is further complicated by the fact that the delivery rate in the hospital doubled within the observation period, and that mothers in different age groups increased in very different proportions.

Their second claim is that there was a significant excess of Group O women among primigravidae. The discrepancy would seem to arise mainly from the extraordinary fact that over the 4-year period, during which the delivery rate doubled, Group A women had more second than first children. Since such conditions would be impossible in a stable or expanding universe there seems no likelihood of this being other than a curious sampling error. The statement that "both for women of English name and for the total Perth sample there is significant excess of Group O women experiencing their first pregnancy" is misleading. The discrepancies of the whole are almost entirely due to the women of English name, who form the majority (for the women of non-English name $\chi^2_{(1)} = 0{\cdot}086$).

The statement, which forms the premise of an argument, that Group O mothers experience a higher number of pregnancies before the age of 19 than Group A mothers, is not supported by the data. The number of women delivered was 139 in Group O and 130 in Group A; expected values are $138{\cdot}4$ and $130{\cdot}6$ respectively.

The analysis of these data derived from a rapidly expanding and highly fertile population is particularly difficult, and does not convincingly demonstrate a primary influence of blood group on fertility.

(5) Johnstone (1954a) published extensive data on mother, husband, and infant phenotype trios from over 2,500 births which occurred in a West London hospital during a 13-month period. The data are

* The word "error" is unfortunate, but not easily replaced. It is merely intended to imply failure to discern the indiscernible.

particularly valuable because they relate to a consecutive series derived from a limited population over a short period, very few mothers therefore being represented more than omce. Husbands as well as mothers were grouped, and many trios had their Rhesus groups tested. Tube methods were used, and the fact that the number of infants of Group AB approximates closely to the expected number and to the number of Group AB mothers suggests that the series was not biassed by systematic grouping errors. Johnstone's analysis of the discrepancies between the expected and observed numbers in husband-wife-child phenotype classes, excluding classes with very small or clearly anomalous frequencies, led to an almost modal value for the discrepancy on the null hypothesis ($\chi^2_{(34)} = 24 \cdot 6$).

A most interesting feature of the data is the large number of children whose blood groups are incompatible with those of their alleged parents. There are no fewer than seventeen bastards* detectable as such. In England the gene frequencies are such that only about 18 per cent. of agamous children can be detected by anomalies in the parent-child ABO blood groups, so that of the 2,578 matings almost a hundred, or nearly 4 per cent., are likely to be agamous.

Efficient estimation of this incidence is difficult. Every trio provides some relevant information, and usually trios with the smallest frequencies are the most informative; in addition most phenotypes contain several genotypes in unknown proportions. The most informative unions are those between recessives; in England about one-fifth of unions are gg/gg (O/O) and, independently, one-thirtieth are dd/dd (rh/rh). Assuming equal phenotype fertility and foetal viability, and the fathers of the agamous children to be almost all different, and to come from the same population, the probability of any paternal gene will be proportional to its frequency in the population. Consider the double recessive marriages, and let a proportion α of the children be agamous, and D be the gene frequency of the dominant gene or genes considered to be known exactly (D = 0·3305 in ABO system, 0·5740 in Rhesus), an assumption which ensures simplicity at a cost of slightly underestimating the variance of the estimate. Let L children be recessive, and B children dominant. Then, to estimate α, the likelihood function will be:

$$B \log \alpha D + L \log (1 - \alpha D).$$

Of the 550 reputed gg/gg (O/O) matings in Johnstone's series, nine resulted in Group A or B children, giving an estimated proportion of agamous children of $4 \cdot 95 \pm 1 \cdot 63$ per cent. Of the 59 reputed dd/dd (Rh/Rh) matings, six resulted in Rh+ children, giving an estimate of $17 \cdot 72 \pm 7 \cdot 07$ per cent. Since very few marriages are likely to be represented in both, only about 0·7 per cent. of marriages being gg,dd/gg,dd and only about 40 per cent. of trios having been Rhesus-typed, these may be combined without gross error, giving an approximate estimate of $5 \cdot 7 \pm 1 \cdot 6$ per cent. This is a remarkably high figure, suggesting that agamous children may even outnumber illegitimate children. Johnstone comments that most of the known cases were first-born. If, as is probable, most of these agamous unions were prenuptial, then since only a minority of women now marry when pregnant, a very large proportion of those who do would appear not to marry the man by whom they are pregnant. If, for example 10 per cent. of first-born children were agamous and all conceived prenuptially, and 20 per cent. of women married while pregnant, half these women would not have married the man by whom they were pregnant.

Johnstone's only claim of significance relates to a shortage of Group O children of O/A marriages. He has pointed out that the data collected by Hirszfeld (1928) and by Waterhouse and Hogben (1947) show a similar shortage, and their data have clearly been approached to test a single prior hypothesis. It would appear that this deficiency of Group O children of O/A marriages is partly attributable to the high incidence of agamous children. This explanation, while sufficient to reduce the level of significance to that conventionally trivial, is not sufficient to allay suspicion of a real effect.

However, Johnstone's analysis of the data of Hirszfeld and of Waterhouse and Hogben is open to two criticisms. Firstly, the two sets of data contain in common the data of Hirszfeld and Hirszfeld and of Hirszfeld, Hirszfeld, and Brokman (Table III, overleaf), comprising 23 O/A marriages with 69 children; the fifteen segregating families contained 21 Group O children and 32 Group A children. This considerable discrepancy has thus been counted twice. Secondly, these data refer not to single births, but to families, and, since there is complete intraclass

* The nomenclature of such children is difficult, bastardy and illegitimacy being legally defined, and conventionally used, in relation to wedlock. It is proposed to term these children, and the union giving rise to them, agamous, as opposed to gamous children whose parents are married or living together; the many illegitimate children born of parents whose union, though stable, is not legal, may conveniently be considered gamous. The words are short, euphonious, and not grossly distorted from their natural meaning. Their occasional use in systematic botany is unlikely to cause confusion.

correlation between the blood groups of all gamous children of gg/G^AG^A marriages, the significance of the discrepancy cannot be assessed by simple tests. An exact solution, with scores, has been presented by Hogben (1932).

In Waterhouse and Hogben's data, the discrepancy is almost entirely due to the non-segregating families, the segregating families containing 272 Group O children and 278 Group A children. The total discrepancy, although its significance is not easily assessed, is suggestive of a high fertility of gg/G^AG^A marriages, a trend which could be a sufficient cause of a persistent variation. If this same discrepancy were evident in Hirszfeld's data, this would be confirmatory; on the other hand, if the discrepancy were of a different nature, the vagaries of sampling, mating, and grouping would appear more likely causes in the apparent absence of any single, sufficient, and reasonable biological explanation. In that part of Hirszfeld's data for which details are available, a considerable discrepancy (about two-thirds of the total) is apparent in those families which segregate. Of the 142 families which contain no obviously agamous children, 72 show segregation, in which there are 131 Group O and 172 Group A children. Bennett and Brandt's efficient test (Table III) gives $\chi^2_{(1)} = 8{\cdot}90$ ($0{\cdot}001 < p < 0{\cdot}01$) with heterogeneity $\chi^2_{(13)} = 18{\cdot}42$ ($0{\cdot}1 < p < 0{\cdot}2$). The ratio of the contributions of the overall effect and the heterogeneity is also significant ($F = 5{\cdot}8$, $0{\cdot}01 < p < 0{\cdot}05$). This does not necessarily indicate that the overall effect is significant, because the greater part of this effect may be due to the presence of agamous children and to grouping errors rather than to the non-systematic disturbances for which analysis by partitioning variance would be appropriate. The data as a whole do not reveal any significant overall effect.

Since the apparently sustained trend in the three sets of data considered by the author is not due to any consistent type of discrepancy, and is not readily accounted for by any hypothesis advanced, and since, where efficient or appropriate tests of significance have been applied on the combined data, there is no unexceptionable evidence of any shortage at even the lowest conventional level of significance, it seems that there is no inexplicable divergence from the null hypothesis of equal fertility and foetal viability of the ABO phenotypes.

(6) Kirk, Shield, Stenhouse, Bryce, and Jakobowicz (1955), in an analysis of a sample similar to, and partially inclusive of, that of Kirk, Kirk, and Stenhouse, drew the following conclusions:

"(*i*) Both in Melbourne and Perth A women had a higher average number of pregnancies than O women, the difference being significant for Perth, where there was also a significant linear increase in this difference with age.

TABLE III

ANALYSIS OF SEGREGATING FAMILIES FROM O/A MARRIAGES IN THE DATA OF HIRSZFELD (1928) AND OF WATERHOUSE AND HOGBEN (1947)

(Method of Bennett and Brandt, 1955)

Authors	No. of Families	No. of Children		Score $\frac{u}{(O-A)}$	Information i	$\frac{\chi^2}{\left(\frac{u^2}{i}\right)}$	d.f.
		Group O	Group A				
Ottenberg	3	7	5	+ 2	6·9524	0·575	1
Learmonth	4	7	11	− 4	12·0000	1·333	1
Pluss	5	11	11	0	14·6513	0·000	1
Kirihara	2	3	5	− 2	4·6667	0·857	1
Hirszfeld, Hirszfeld, and Brokman	5	8	10	− 2	9·3335	0·429	1
Mino	3	4	6	− 2	5·0323	0·795	1
Staquet	8	19	27	− 8	36·6989	1·744	1
Furuhata	3	6	3	+ 3	3·0000	3·000	1
Snyder	20	34	57	−23	61·5408	8·596	1
Heim	2	2	3	− 1	1·0000	1·000	1
Bais and Verhoef	4	7	7	0	6·5714	0·000	1
Schiff	3	10	5	+ 5	10·9847	2·276	1
Hirszfeld and Hirszfeld	10	13	22	− 9	16·5238	4·902	1
						25·507	13
Total of Above (Hirszfeld, 1928)	72	131	172	−41	188·9558	8·896	1
Waterhouse and Hogben (1947)*	102	249	247	+ 2	306·4970	7·681	11
Grand Total	174	380	419	−39	495·4528	33·188	24
						3·070	1

* Less the fifteen families already included in Hirszfeld's series.
From the analysis of Bennett and Brandt (1955).

"(*ii*) A comparison of the calculated and observed number of children in different mother-child blood-group combinations for 16,179 single births in Melbourne reveals a significant deficiency of Group A children to Group O mothers and Group AB children to Group A and B mothers, and a significant excess of Group O children to Group O mothers, of Group A children to Group B mothers, and Group B children to Group A mothers.

"(*iii*) A breakdown of mother-child blood-group combinations by birth rank reveals that the deficiency in Group A and B children to Group O mothers is absent for first-born children, present for second-born children, and greatest for third-born children."

Their statement of the significant excess of children to Group A mothers is based on a "*t*" test which is probably not appropriate, as the variance of the number of children per family, especially when the first is excluded, as a necessary condition of motherhood, usually greatly exceeds the mean. No details of the linear increase with age are given, but it is apparently based on the number of children born by a given age and not the number being born at that age.

Their second set of claims, although maintained at very high levels of significance, are not necessarily acceptable evidence of differing fertility, since their comparison of phenotype frequencies of mother and child ($\chi^2_{(3)} = 23{\cdot}22$) suggests effects of such magnitude that they would have shown up in population surveys covering different generations and sexes. It seems more likely that there were systematic grouping errors. This explanation is supported by the discrepancy in blood group distribution between mothers and infants ($\chi^2_{(3)} = 31{\cdot}87$). If this discrepancy is examined in detail (Table IV), it is evident that the greater part is due to a shortage of Group AB infants. If Groups B and AB are combined, an efficient estimate of the proportionate gene frequencies may be made by solving the following determinate equations, where O, A, B, and AB are the proportionate phenotype frequencies:

$$r^2 = \mathrm{O}$$
$$2pr + p^2 = \mathrm{A}$$
$$2pq + 2qr + q^2 = \mathrm{AB} + \mathrm{B}$$

These are satisfied by the values $p = 0{\cdot}2420$, $q = 0{\cdot}0658$, and $r = 0{\cdot}6922$, giving the expected proportion of AB as 3·185 per cent., the observed proportion being 2·695 per cent., which implies that about 15·4 per cent. of Group AB infants were wrongly grouped as B. This estimate is similar to that obtained from Bryce and Jakobowicz's former data (11·6 ± 4·5 per cent.). After corrrecting for grouping errors by the use of this estimate, comparison of the phenotype distributions of mothers and children shows much smaller discrepancies (Table IV).

TABLE IV
DISTRIBUTION OF MOTHERS AND CHILDREN ACCORDING TO BLOOD GROUP

Blood Group	No. of Mothers	No. of Children	χ^2	χ^2 (corrected*)
O	7,585	7,752	1·818	1·818
A	6,406	6,369	0·107	0·107
B	1,605	1,622	0·090	1·221
AB	583	436	21·206	4·211
Total	16,179	16,179	23·221	7·357

When $\chi^2_{(3)} = 23{\cdot}221$, $p < 0{\cdot}001$
$\chi^2_{(3)} = 7{\cdot}357$, $0{\cdot}01 < p < 0{\cdot}05$
* Corrected for estimated grouping errors.

The discrepancies remaining are due to an excess of children of Group O and a deficiency of Groups A, B, and AB. This is not unexpected, since the data relate to at least two racial sub-groups with different gene frequencies, and since mating is assortative (as is demonstrated by the different gene frequencies of mothers with and without English names) the children would be expected to contain an excess of homozygous individuals over those calculated on the assumption of random mating. For the same reason gene frequencies calculated as above, on the assumption of random mating (or random pairing of allelomorphs), and any estimate based on this calculation, must be regarded as approximate. Both assortative mating and grouping errors would lead to an apparent deficiency in the observed numbers of Group AB infants.

Their third claim regarding the influence of birth rank on infant blood group cannot be regarded as clearly demonstrated if consideration is extended to the higher birth ranks. The relevant data (Table V, overleaf), although showing an apparent loss of incompatible children in the second and third birth ranks, do not show any consistent trend when viewed as a whole.

(7) Matsunaga (1953) and Matsunaga and Itoh (1953) in Japan, have made the only large surveys on entire families within defined districts. Since the districts were uniform for race and the families were large, these data would be of extreme value, but unfortunately only a part has so far been published. This shows an extreme impairment in fertility

TABLE V
FREQUENCY OF INCOMPATIBLE CHILDREN* ACCORDING TO BIRTH RANK

Birth Rank	No. of Children			Percentage of Incompatible Children
	Compatible	Incompatible	Total	
1	2,113	488	2,601	18·8
2	1,930	436	2,366	18·4
3	1,221	233	1,454	16·0
4	639	145	784	18·5
5 and Over	690	160	850	18·8
Total	6,593	1,462	8,055	18·2

* An incompatible child is one who would be incompatible (relative to the ABO system) as a blood donor to his mother.

and infant viability within heterospecific marriages, an effect which, it is suggested, may be balanced by mutation.

Some of their data are more widely known (Matsunaga, 1955) and are said to imply a significant deficiency of Group A children of Group O mothers and Group A fathers, and of Group B children of Group O mothers and Group B fathers. However, the significance has not yet been rigorously demonstrated, since the tests assume the proportion of homozygous fathers to have zero sampling variance, the distribution of family size to be Poissonian, and husbands to be synonymous with fathers. The apparent increase in abortion rate in heterospecific marriages assumes that a woman who has had one abortion is not more likely to have another. The data are such that no reasonable interpretation or informed criticism can precede complete publication.

A Critical Examination of Published Work claiming an Association between Blood Group and Sex-Ratio

The influence of parental blood groups on the sex-ratio of children, which may be regarded as a particular form of differential fertility, has been the subject of three widely-quoted reports:

(1) Sanghvi (1951) claimed consistent significant discrepancies on two mother-child series, one of 1,330 Indian births, the other of 865 births in New York. When each series is analysed as a whole, the data, excluding phenotype pairs of low frequency, do not show any significant heterogeneity ($\chi^2_{(13)} = 13{\cdot}9$ for the Indian series and $\chi^2_{(6)} = 9{\cdot}9$ for the New York series).

However, Sanghvi has not made this analysis, but has analysed only those classes in the body of the table which are "large enough for further consideration". If these classes were selected purely on grounds of size, the chance that the difference between the outlying pair would be significant at the p^{th} level considerably exceeds p. With such large numbers, in any class frequency the probability distribution of $\frac{d}{\sqrt{e}}$ will be approximately that of a random sample from a normal variate of unit variance, and the range of the extremes in such cases has been tabulated (May, 1952). The range of three such samples which will just not be exceeded in 5 per cent. of cases is 3·32 units, and a $2 \times 2\ \chi^2$ estimation will be at least $2(\frac{1}{2} \times 3{\cdot}32)^2$ or 5·51, in 5 per cent. of cases. Sanghvi's value of $\chi^2_{(1)} = 5{\cdot}57$ is possibly just significant at this level. The New York data where then examined in the light of this prior hypothesis, which was confirmed, ($\chi^2_{(1)} = 4{\cdot}93$), although the high overall sex-ratio possibly impairs any assumption of generality.

This appears to be the only rigorously demonstrated discrepancy in relation to sex-ratio and blood group. However, in view of the failure of more recent work to confirm these findings, it appears likely that this promising start was due to the fact that when tests are used to assess significance at the 5 per cent. level, 5 per cent. of null hypotheses will inevitably be wrongly rejected through chance alone.

(2) Johnstone (1954b) published data referring to 3,791 children of 2,100 marriages, the sexes of the children and the blood groups of both parents being known. This is by far the most informative series yet published. Having noted that the sex-ratio of infants of Group B mothers was high, while that of Group B fathers was low, he picked out for special attention the two marriage classes B/not-B and not-B/B and found a difference in the sex-ratio of the children which almost reached the conventional threshold of significance ($\chi^2_{(1)} = 3{\cdot}799$). Little importance can be attached to this result, however, since numerous equally reasonable comparisons could have been made. For example, if fourteen appropriate independent comparisons could be made, there would usually be at least one "statistically significant" discrepancy, even if there were no association between sex-ratio and blood group.

A further analysis of the B/not-B marriages by birth rank has inevitably led to discrepancies which appear to be even more significant. As there are several possible comparisons, the highly significant discrepancy reputed to occur between the sex-ratios of first children of the reciprocal marriage pairs cannot be accepted; and since it is a posterior hypothesis unsupported by exterior evidence or by any explanatory mechanism it must provisionally be considered as a vagary of sampling. The data as a whole, excluding three phenotype pairs with less than ten

children of either sex, do not reveal any significant heterogeneity ($\chi^2_{(18)} = 15{\cdot}9$).

(3) Cohen and Glass (1956), in an analysis of an extensive series of mother-child phenotype pairs assembled in Baltimore, were unable to demonstrate any significant association with sex-ratio. They pointed out that Johnstone's data showed no discrepancy when similarly analysed, and this absence of any significant discrepancy persisted after his data were combined with theirs. But the addition of Sanghvi's data from New York and Bombay enabled them to demonstrate highly significant discrepancies in a large number of comparisons; in particular, the sex-ratio was low in offspring of Group A mothers and in Group A offspring, and was high in offspring of Group B mothers and in Group B offspring.

The discrepancies found are not other than those to be expected if data from Baltimore and Isleworth are combined with data from Bombay, since in the Indian data the sex-ratio was very high (54·6 per cent.), and the frequency of Group B relative to Group A much higher than in the British and American data. Consequently the infants of Group B mothers and the Group B infants contain a disproportionate number of Indians, who are poorly represented among both infants of Group A mothers and Group A infants. The data of Sanghvi are said not to differ significantly from those to which they were added ($\chi^2_{(28)} = 25{\cdot}44$), but as a similar test for homogeneity of the Baltimore sample gave the almost impossible result $\chi^2_{(14)} = 0{\cdot}67\,(p > 0{\cdot}999999)$, this claim of homogeneity cannot be accepted. As the sex-ratio and the proportions of the various blood groups in the Indian and not-Indian data differ significantly and considerably, to an extent and in a direction sufficient to explain the overall heterogeneity, the analysis of the data cannot be accepted as necessarily implying any primary relationship of ABO phenotype and sex-ratio.

Summary

Evidence is advanced that grouping errors (over 10 per cent. of Group AB infants being grouped as B in two series) and paternity errors (over 5 per cent. of children being not fathered by their mother's husband in one series) are so common as greatly to complicate any analysis of data suggestive of a relationship between fertility and blood group.

It is concluded, for these and other reasons, that there has not yet been any satisfactory demonstration at the statistical level of any differences in fertility or sex-ratio primarily related to the ABO blood groups.

The apparent absence of any statistical evidence is probably quite compatible with differences in fertility sufficient to maintain the variation in the ABO phenotypes.

References

Allan, T. M. (1953). *British Journal of Preventive and Social Medicine*, **7**, 220.
Bennett, J. H. and Brandt, J. (1954). *Ann. Eugen., (Camb.)* **18**, 302.
British Medical Journal, **1**, 1197 (Annotation).
Bryce, L. M., Jacobowicz, R., McArthur, N., and Penrose, L. S. (1950). *Ann. Eugen. (Camb.)*, **15**, 271.
Cohen, B. H., and Glass, B. (1956). *Hum. Biol.*, **28**, 20.
Glass, B. (1950). *Amer. J. hum. Genet.*, **2**, 269.
Haldane, J. B. S. (1956). *Ann. hum. Genet.*, **20**, 309.
Hirszfeld, L. (1928). "Konstitutionsserologie und Blutgruppen-Forschung". Springer, Berlin.
Hogben, L. (1932). *J. Genet.*, **25**, 211.
Huxley, J. (1955). *Heredity*, **9**, 1.
Kirk, R. L., Kirk, M., and Stenhouse, N. S. (1953). *British Journal of Preventive and Social Medicine*, **7**, 1.
Kirk, R. L., Shield, J. W., Stenhouse, N. S., Bryce, L. M., and Jacobowicz, R. (1955). *Ibid.*, **9**, 104.
Johnstone, J. M. (1954a). *Ibid.*, **8**, 117.
—— (1954b). *Ibid.*, **8**, 124.
Matsunaga, E. (1953). *Proc. Jap. Acad.*, **29**, 399.
—— (1955). *Amer. J. hum. Genet.*, **7**, 66.
—— and Itoh, S. (1953). *Proc. Jap. Acad.*, **29**, 529.
May, J. M. (1952). *Biometrika*, **39**, 192.
Penrose, L. S. (1955). Personal Communication.
Sanghvi, L. D. (1951). *Nature (Lond.)*, **168**, 1077.
Waterhouse, J. A. H., and Hogben, L. (1947). *British Journal of Social Medicine*, **1**, 1.
Woolf, B. (1955). *Ann. Eugen. (Camb.)*, **19**, 251.

Brit. J. prev. soc. Med. (1957), **11**, 90–101

RECENT TRENDS OF MORTALITY FROM VIOLENCE

BY

W. J. MARTIN

Statistical Research Unit, of the Medical Research Council, London School of Hygiene and Tropical Medicine.

The causes of violent deaths have changed with the growth of civilization. In England the risk of being murdered or hanged is much smaller than three centuries ago when life was held to be cheaper, but in the last 150 years new and potent methods of killing have been introduced. With the development of industrialization new hazards arose from machinery, from railways, and from the internal combustion engine. In the last decades of the 19th century, there was an increasing awareness of the unnecessary waste of life from accidents in industry. The various risks in industry have been greatly reduced, but the more recent campaign against road accidents, although continuous and vigorous, has not met with any great success.

As a result of the measures taken, the death rate from accidents fell dramatically. In males, the death rate from accidents and negligence fell from 1,064 per million in 1861–70 to 496 in 1931–35, and in females it declined from 314 to 198 during this period. The death rate from accidents in coal mines fell from 96 per million in 1863–70 to forty in 1931–38. On the railways the death rate in males of 89 per million in 1871–80 fell to only twenty in 1931–38. In males, deaths from accidental drowning were 214 per million in 1863–70 and 58 in 1931–38, and during this period the death rate in males from accidental burns and scalds fell from 109 to 29 per million. The death rate in males from all forms of suffocation was 73 in 1863–70, increased to 81 in 1891–1900, and then fell steadily to twenty in 1931–38. The death rate from accidents due to horse vehicles was 124 per million in 1863–70, declined to 59 in 1901–10, and fell to five in 1931–38.

The very large fall in fatal accidents due to horse vehicles has been more than compensated for, however, by motor vehicles. The first deaths registered as due to motor cars were one male and one female in 1896 and the first motor-cycle deaths registered were two males in 1902. From these small beginnings a terrific stream of deaths has resulted; in 1954 the numbers killed by motor vehicles were 3,289 males and 1,158 females, including 1,049 males and seventy females who died from motor-cycle accidents.

Aircraft have begun to be an important cause of violent death. In 1911 six deaths were registered, in 1921 the numbers rose to 64, including 38 in the disaster to airship R38, and in 1954 the number of deaths was 225. The largest annual number of deaths from aircraft accidents was 311 in 1952.

The introduction of gas for domestic purposes provided another means of suicide that has become more frequent during the present century.

Although there was a large fall in the mortality from violence before 1939, the rate of decline has been much slower than that of the death rate from natural causes. As a result of this trend violence has become relatively more important as a cause of death; in 1861–71 deaths from violence formed 3·4 per cent. of all deaths and in 1931–39 the proportion was 4·6 per cent.

During the past 20 years there have been changes in the mortality from violence. The chief is the decline in mortality in males; in females the death rate has fluctuated about a fairly constant level. In 1935 the death rate per million was 743 for males and 345 for females. and in 1954 the rates were 593 and 358 respectively. The actual number of deaths was 14,479 males and 7,301 females in 1935 and 12,630 males and 8,239 females in 1954.

The standardized death rates from violence during the past 90 years are shown in Table I. The death rates from all forms of violence have declined faster for males than for females. In males the rate in recent years is less than half the rate at the beginning of the period, while in females it is about two-thirds of its former level; consequently the death rate in males is now 2½ times that in females compared with 3½ times at the beginning of the period. The course of the standardized death rate for suicide has been

different for the two sexes. In males the rate has fluctuated considerably, but in recent years has been slightly below the most favourable rates of the earlier years. In females, on the other hand, the rate increased from 1861 to 1940, fell during the years 1941–45, and has since risen slightly again to a level above that of the years 1861–1900. A fall in the suicide rate was recorded for both sexes during the two world wars, 1914–18 and 1940–45.

TABLE I

STANDARDIZED RATES PER MILLION FROM VIOLENCE (EXCLUDING WAR CASUALTIES) IN ENGLAND AND WALES, 1861–1954

Period	Suicide		Other Violence		All Violence	
	Males	Females	Males	Females	Males	Females
1861–70	105	35	1,064	314	1,169	349
1871–80	115	36	993	304	1,108	340
1881–90	128	39	829	284	957	323
1891–00	144	44	790	313	934	357
1901–10	157	47	670	282	827	329
1911–20	119	43	729	255	848	298
1921–30	138	52	523	201	661	253
1931–40	140	60	519	264	659	324
1941	96	44	702	254	798	298
1942	89	43	585	224	674	267
1943	95	44	530	200	625	244
1944	97	40	520	206	617	246
1945	95	46	506	202	601	248
1946	99	51	431	182	530	233
1947	95	51	447	184	542	235
1948	101	52	370	161	471	213
1949	101	50	383	155	484	205
1950	96	46	389	152	485	198
1951	94	46	423	156	517	202
1952	93	43	416	147	509	190
1953	101	48	402	153	503	201
1954	105	52	394	154	499	206

ACCIDENTS

Table II shows the age distribution of death rates from violence, excluding suicide and war casualities, by age and sex for the years 1941 to 1954. The largest improvement in both sexes has been at ages 5 to 9, where the death rate has fallen to one-third of the level of 1941. The death rates from birth to 4 years and from 10 to 14 years have fallen to less than one-half of their initial level. The improvement for the age groups under 15 was similar for both sexes; at ages 25–74 the rate in males has been halved, while that in females has fallen by approximately one-third. The smallest rate of decline was in old age, 75 and over. Excluding old age, the smallest rate of improvement among males was at ages 15–19 and 20–24, the rates in 1954 being 72 and 82 per cent. respectively of the rates in 1941. In childhood, the rate fell steadily during the period, but at older ages it declined to a minimum in the middle of the period with a tendency to rise in the more recent years. This happened in both sexes, but was more marked in males, especially in the age groups 15 to 40, where the minimum was reached in 1948. At ages 20–24 the death rate fell from 741 per million in 1941 to 344 in 1948, rose to 684 in 1952, and declined to 608 in 1954.

TABLE II

DEATH RATES PER MILLION FROM VIOLENCE, (EXCLUDING SUICIDE AND WAR CASUALTIES) ENGLAND AND WALES, 1941–54

Sex	Year	Age Group (yrs)										
		0–	5–	10–	15–	20–	25–	35–	45–	55–	65–	75 and Over
Males	1941	839	523	348	553	741	648	584	696	1,000	1,403	3,168
	1942	790	486	277	450	637	508	459	560	768	1,115	2,747
	1943	720	481	263	417	546	480	418	496	651	946	2,333
	1944	767	494	276	404	538	438	424	454	645	961	2,188
	1945	736	498	297	459	623	393	348	388	556	863	2,204
	1946	688	328	242	389	461	312	310	371	538	810	2,146
	1947	664	381	225	388	476	343	332	414	517	815	2,304
	1948	585	318	176	318	344	285	267	343	484	658	1,903
	1949	547	299	193	383	438	284	294	338	450	650	2,063
	1950	461	252	151	399	531	360	296	350	463	679	2,021
	1951	489	261	182	403	639	406	308	368	483	707	2,272
	1952	472	219	163	455	684	379	318	348	471	706	2,044
	1953	418	217	149	409	641	366	303	360	469	771	2,332
	1954	393	168	158	399	608	339	303	349	484	809	2,772
Females	1941	623	247	105	116	97	82	90	144	283	793	3,253
	1942	564	231	86	88	106	86	86	118	222	617	2,851
	1943	521	180	98	82	79	72	75	102	186	533	2,582
	1944	566	224	89	76	72	72	68	104	210	528	2,513
	1945	561	209	94	78	74	71	69	98	187	537	2,455
	1946	494	149	69	67	54	63	63	90	193	514	2,634
	1947	503	162	63	70	54	57	65	103	195	536	2,593
	1948	434	153	63	60	56	48	57	89	163	440	2,243
	1949	387	128	63	67	65	40	51	85	171	452	2,374
	1950	338	127	46	70	58	45	50	100	166	454	2,582
	1951	350	97	45	78	68	47	59	94	167	481	2,743
	1952	330	101	49	66	75	50	53	95	169	441	2,344
	1953	318	95	59	62	65	50	60	105	183	501	2,636
	1954	264	86	48	70	67	55	60	105	190	586	2,936

The numerically important kinds of accidents for the years 1941–54 are shown in Table III. During the 14 years a substantial decline occurred in all forms of fatal accidents tabulated, except those from suffocation and poisoning, which have increased slightly, and those in females from falls. The death rate from motor vehicle accidents declined to a minimum in 1948 for males and in 1949 for females and then rose. As a result, of this rise the mortality has reverted to the level of 9 years earlier. Deaths from falls also show a decline to a minimum in 1948 and a subsequent rise, but though the rates in males remained considerably below the initial level, in females the rate in 1954 was slightly above its 1941 level. The chance of a fatal accident from burns was the same for males and females in the early 1940s, but has fallen faster for males than for females; the chance of dying from accidental burns in 1954 was twice as large for females as for males. The mortality from accidental drowning has fallen considerably in males but has shown only a slight improvement in females. As a result the ratio has fallen, from 7 to 1 in 1941 to 4 to 1 in 1954.

The death rates from the more important accidental causes are shown in Table IV by age and sex for three quinquennial periods. In males the death rates from accidental drowning have decreased by one-half in each age group under 65; in females the fall has been mainly confined to ages under 10. The age trend of the rates is different in the two sexes. In each of the three quinquennia, the rate in males is high among pre-school children, rises at ages 5–9, falls at ages 10–14, and rises again at ages 15–19. With the exception of ages 75 and over, the chance of a drowning accident is highest at ages 5–9. In females, the death rate was largest at ages under 5 years, fell to a constant level in early middle life, and rose at the older ages.

TABLE III

DEATH RATES PER MILLION FROM ACCIDENTS, BY CAUSE, AND SEX, ENGLAND AND WALES, 1941–1954

Year	Railways		Motor Vehicle		Poisoned by Solid and Liquid Substances		Poisoned by Gases and Vapours		Falls		Fire and Explosion of Combustible Material		Suffocation		Drowning		Others		All Accidents	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
1941	26	2	281	74	5	4	12	6	145	136	44	43	24	15	75	11	159	20	771	311
1942	26	2	208	56	5	3	9	6	127	120	41	39	24	13	59	10	139	22	638	271
1943	27	3	165	46	4	2	8	6	118	114	32	32	24	11	60	11	139	22	577	247
1944	28	3	178	55	4	3	10	7	107	111	31	32	27	16	63	10	129	18	577	255
1945	25	3	158	49	5	4	13	9	103	110	27	31	29	15	65	11	113	21	538	253
1946	22	3	153	47	5	4	10	10	88	110	20	27	30	18	44	11	96	15	468	245
1947	23	2	147	47	5	5	11	12	94	110	27	27	34	21	50	10	92	18	483	252
1948	17	3	128	44	6	4	10	9	78	100	13	20	26	17	39	11	85	14	402	222
1949	16	1	141	41	5	4	11	10	81	105	9	16	25	16	40	8	90	21	418	222
1950	16	2	151	46	5	6	10	11	74	113	9	16	23	14	36	9	97	16	421	233
1951	14	1	161	49	6	6	12	13	86	117	15	19	23	14	36	9	98	17	451	245
1952	18	3	149	42	7	6	12	13	79	105	13	19	22	13	34	8	97	17	431	226
1953	16	1	158	45	6	7	15	14	84	123	10	16	19	13	34	8	94	22	436	249
1954	12	1	161	51	7	7	15	17	99	141	10	19	23	13	30	8	92	15	449	272

DEATH RATES PER MILLION FROM CERTAIN ACCIDE

Type of Accident	Quinquennium	0–		5–		10–		15–		20–	
		Male	Female	Male	Female	Male	Female	Male	Female	Male	Fe
Drowning	1940–44	92	32	120	14	64	9	69	4	71	
	1945–49	75	24	96	13	53	11	68	7	41	
	1950–54	49	18	63	8	31	6	42	4	29	
Falls	1940–44	38	24	35	13	26	7	46	4	62	
	1945–49	28	16	25	7	22	5	32	5	36	
	1950–54	17	12	12	4	14	2	24	2	32	
Fire	1940–44	155	135	25	53	8	20	12	13	12	
	1945–49	61	70	11	31	5	9	5	4	2	
	1950–54	13	28	6	19	3	7	3	4	5	
Motor Vehicles	1940–44	145	92	246	119	114	41	161	43	241	
	1945–49	112	64	182	90	87	32	145	32	210	
	1950–54	90	52	104	54	64	21	201	38	337	

During the 15 years fatal falls have declined in every age group except in old age (75 and over), where the rates in 1950–54 exceeded those in 1945–49. The rate in males was considerably in excess of that in females in each age group up to age 65. The death rate in females slightly exceeded that in males at ages 65–70 and was considerably above it at ages 75 and over. Since 1949, the place of occurrence has been tabulated, and the percentage distribution of 8,974 fatal falls in males and 13,657 in females during 1950–54 is shown in Table V. The differences between the sexes are partly due to the difference in occupational risk. Fatal falls from one level to another occurred at work in 50·4 per cent. of accidents in males and in only 1·1 per cent, in females.

TABLE V
DEATHS FROM FALLS, BY SEX

Place of Occurrence	Percentage of Fatal Falls	
	Males	Females
Stairs	17·9	18·3
Ladders	5·0	0·3
One Level to Another ..	29·9	11·3
Same Level	28·7	39·8
Unspecified	18·5	30·4
Total	100	100

The number of fatal falls at home by age and sex is shown in Table VI for the quinquennium 1950–54. There is a considerable difference between the sexes; in each age group females have a larger proportion of fatal accidents at home than males, but the actual number of fatal accidents at home is larger for males in each age group up to age 45. At ages over 65 females have rather more than 2½ times as many fatal falls at home as males. The proportion of fatal falls that occurred at home at ages 65 and over was 81 per cent. for males and 93 per cent. for females.

TABLE VI
DEATHS FROM ACCIDENTAL FALLS, BY AGE AND SEX, ENGLAND AND WALES, 1950–54

Sex		Males			Females		
		Total Number of Falls	Falls at Home or in Residential Institutions		Total Number of Falls	Falls at Home or in Residential Institutions	
			No.	Per Cent.		No.	Per Cent.
Age	0–	155	117	75·5	103	82	79·6
	5–	208	41	19·7	54	14	25·9
	15–	1,381	242	17·5	164	98	59·8
	45–	1,568	571	36·4	799	619	77·5
	65 and Over	5,662	4,240	74·9	12,537	10,835	86·4
	Total	8,974	5,211	58·1	13,657	11,648	85·3

The risk of a fatal burn (Table IV) has fallen during the 15 years in each age group in both sexes, the greatest improvement being in childhood. The risk of a fatal burn is much larger among females under 10 years of age and over 65 years than among males, and the fall in the death rate in childhood has been greater among boys than among girls. At ages 0–4 the rate in girls was slightly below that in boys in 1940–44; in 1950–54, however, it was over twice that in boys. At ages 5–9 the rate in girls was higher than in boys; in a ratio rising from 2 to 3 in the 15 years.

In each age group in both sexes the death rate from motor vehicle accidents (Table IV) was lower in 1945–49 than in 1940–44, but rose again in 1950–54 in each male age group over age 15. This rise was most marked in the age groups 15–19, 20–24, and 25–29; in the first two age groups the male rate for 1950–54 was considerably above the rate for 1940–44.

GE AND SEX, ENGLAND AND WALES, 1940–1954

(yrs)

25–		35–		45–		55–		65–		75 and Over		Total	
Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
6		37	6	49	10	64	13	77	15	80	14	64	11
4		19	6	28	10	41	13	65	15	86	13	46	10
3		17	5	24	9	37	11	62	17	93	17	35	9
6		69	12	101	30	179	94	383	398	1,608	2,340	139	129
4		45	5	58	15	98	57	232	261	1,228	1,887	93	107
3		36	5	51	13	79	46	235	245	1,305	2,029	87	120
9		9	9	10	16	18	31	48	77	160	254	32	38
5		5	4	6	9	12	19	27	48	115	210	16	25
2		8	3	9	6	13	13	21	40	109	161	12	18
30		154	27	180	41	271	66	411	120	662	188	215	59
17		95	20	103	27	146	52	234	97	443	165	142	45
19		109	20	110	32	151	51	234	98	484	194	154	46

Deaths of motor cyclists contributed largely to the excess in these age groups. The distribution by type of accident in 1950–54 for ages 15–19 and 20–24 is shown in Table VII. The absence of the numbers exposed to risk in these eight groups makes impossible any inferences about the relative risks of different types of accident. For example, nothing can be inferred about the relative danger of pedal cycles in the two age groups, although the number of pedal cyclists killed at ages 15–19 was twice the number at ages 20–24; the younger age group may contain many more users of pedal cycles than the older, where, perhaps, pedal cycles have been replaced by motor cycles or motor cars.

TABLE VII

DEATHS OF YOUNG ADULT MALES FROM ROAD ACCIDENTS

Road Users	Males			
	Ages 15–19		Ages 20–24	
	No.	Per Cent.	No.	Per Cent.
Pedestrians	108	9·4	120	5·8
Pedal Cyclists	280	24·5	136	6·6
Riders or Passengers on Motor Cycles	680	53·1	1,516	73·2
All Others	149	13·0	300	14·4
Total	1,145	100·0	2,072	100·0

Lack of adult supervision plays an important part in accidental deaths among pre-school children. The variation in the death rates from different types of accidents during the first years of life are shown in Table VIII for the years 1950–54. Boys had a higher death rate at each age and for every cause than girls, except for burns and scalds. In boys, the death rate from burns and scalds rose to a maximum at age 1 and fell steadily to one-quarter of this level by age 4–5. In girls, it was at a maximum at age 2 and then fell slowly, so that the death rate of girls at age 4–5 was three times that of boys. The death rate from drowning was highest at age 2, its subsequent fall being more rapid for girls than for boys. The death rate from falls was fairly constant during the first 4 years of life and then fell by one-half in the fifth year. In boys, the risk of a fatal motor vehicle accident increased throughout the first five years of life; in girls, it fell slightly at ages 4–5. Of the deaths from suffocation during the first year of life, 50 per cent. were due to inhalation or ingestion of food, and 46 per cent. to suffocation in cot or cradle. Accidental poisoning arises from a wide variety of substances. Swinscow (1953), in a detailed analysis of deaths from accidental poisoning among children aged 1–4 during 1931–49, found that slightly over one-half of the deaths were due to medicaments and about one-quarter to household poisons. About one quarter of the medicaments were liniments of various kinds, and the remainder were mainly preparations which were regarded as harmless to adults even in large doses. Household poisons included rat poison, insecticides, turpentine, etc., but excluded disinfectants, which accounted for about 7 per cent. of the deaths.

SUICIDE

The death rates from suicide by age and sex for recent years are shown in Table IX (opposite). In males, the suicide rate steadily increased throughout life, but, in females, it fell sharply at ages 75 and over to below the rates for ages 45–64 and 65–74. The largest difference between the death rates of the two sexes is seen at ages 75 and over. During the past 14 years changes have occurred in the age and sex distribution of these deaths. The suicide rate has declined for both sexes under the age of 25 years but has increased at ages 45 and over. At ages 15–19 and 20–24 this fall has been less for females than for males. At the beginning of the period, the rate in males in these age groups was 3 to 4 times greater than that in females; in the last years it was 2½ to 3 times greater. At the other extreme of life, the rate has risen faster in females than in males. At ages 65–74, the ratio fell from 3 to 2 during the period. At ages 75 and over, the rate in males was from 5 to 6 times that in females at the beginning of the period and about 3½ times at the end.

TABLE VIII

DEATH RATE PER MILLION FROM ACCIDENTS AMONG CHILDREN UNDER 5 IN ENGLAND AND WALES, BY SEX, 1950–54

Cause	Age (yrs)									
	0–		1–		2–		3–		4–5	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Scalds and Burns	28	29	52	47	33	51	27	48	14	42
Drowning	7	7	44	22	73	30	67	19	53	13
Falling, Crushing	23	13	26	16	23	17	20	13	13	6
Motor Vehicles	8	4	61	40	117	73	124	76	132	64
Suffocation	858	641	49	35	11	8	4	4	3	1
Poisoning	8	7	43	28	14	10	12	4	4	6

TABLE IX

DEATH RATES PER MILLION FROM SUICIDE, BY AGE AND SEX, ENGLAND AND WALES, 1941–1954

Sex	Year	Age Group (yrs)								
		10–	15–	20–	25–	35–	45–	65–	75 and Over	All Ages
Males	1941	3	37	81	100	130	232	374	414	135
	1942	3	42	77	96	117	211	331	366	125
	1943	2	38	104	95	140	219	348	314	134
	1944	3	34	111	108	126	226	332	395	135
	1945	4	34	82	110	116	226	352	421	136
	1946	5	30	71	93	145	244	391	465	144
	1947	3	36	69	85	114	254	383	480	137
	1948	2	29	76	83	131	269	469	388	145
	1949	1	33	62	81	134	278	422	490	147
	1950	1	35	64	72	124	265	416	421	136
	1951	6	29	61	80	120	249	412	478	134
	1952	1	40	64	81	120	247	390	414	132
	1953	1	33	80	91	127	263	412	482	142
	1954	3	30	69	96	147	278	430	439	149
Females	1941	0	11	20	55	70	126	117	67	62
	1942	1	11	18	48	78	118	114	81	62
	1943	1	7	25	54	86	112	117	66	63
	1944	0	5	24	48	74	107	114	70	58
	1945	1	10	27	55	78	123	122	79	66
	1946	1	15	27	52	87	145	146	92	75
	1947	0	11	28	51	80	146	166	114	76
	1948	0	11	20	50	80	160	173	98	79
	1949	1	15	26	45	77	144	165	138	75
	1950	1	10	23	35	75	139	153	115	70
	1951	0	9	20	38	66	146	167	107	72
	1952	1	11	12	36	66	134	165	98	68
	1953	3	10	22	39	79	145	171	128	76
	1954	0	12	23	52	77	149	198	130	81

Table X shows that the relative frequencies of different methods of suicide has also changed during recent years. The use of analgesic and soporific drugs has increased and was about four times as common at the end of the period as at the beginning. This method of suicide is the only one in which the number of females exceeded the number of males. The use of coal gas has steadily increased, the death rates having risen by 50 per cent. for both sexes during the 14 years. The use of other poisons has

TABLE X

DEATH RATES PER MILLION FROM SUICIDE, BY METHOD AND SEX, ENGLAND AND WALES, 1941–1954

Sex	Year	Poisoning			Hanging and Strangulation	Drowning	Firearms and Explosives	Cutting and Piercing	Jumping from High Places	Other Methods
		Analgesic and Soporific Substances	Domestic Gas	Other Poisons						
Males	1941	3	42	7	24	17	15	16	4	6
	1942	2	38	6	24	16	14	13	5	6
	1943	2	41	7	27	18	14	15	4	5
	1944	2	46	7	24	19	12	15	4	6
	1945	3	45	7	29	18	11	14	3	5
	1946	3	46	7	32	18	11	15	4	7
	1947	5	47	6	30	15	12	11	4	7
	1948	7	54	6	29	17	10	9	5	7
	1949	7	53	6	31	17	11	11	4	6
	1950	8	49	5	27	16	9	9	4	8
	1951	7	51	5	25	15	9	8	4	7
	1952	11	52	5	25	13	8	7	3	7
	1953	9	60	5	25	12	10	9	3	7
	1954	10	64	4	29	14	9	8	4	7
Females	1941	3	28	7	6	11	1	2	2	3
	1942	3	29	6	5	11	1	2	3	2
	1943	3	29	5	6	13	1	3	2	2
	1944	3	28	5	5	11	1	2	2	2
	1945	4	31	5	6	12	1	2	3	2
	1946	5	36	5	8	12	0	3	3	3
	1947	6	39	5	7	11	1	2	3	2
	1948	10	40	3	6	11	1	2	3	3
	1949	9	38	4	6	10	0	2	3	2
	1950	11	33	4	6	10	0	2	2	2
	1951	10	39	3	6	10	1	1	2	2
	1952	13	34	3	5	8	0	1	2	1
	1953	11	41	3	6	9	0	1	2	1
	1954	14	47	2	6	8	0	1	2	2

fallen by half. The death rate from hanging has risen slightly for males but has remained fairly constant for females. Other methods of suicide—drowning, firearms, cutting and piercing, jumping from high places—were less frequent at the end than at the beginning of the period.

The distribution of the methods used was quite different for the two sexes, although coal gas poisoning was the most frequent means of suicide for both sexes. The distribution of deaths by method of suicide for 1954 is shown in Table XI. Poisoning by coal gas has been numerically the most important method of suicide for the past 20 years. During this period the number of deaths has increased more rapidly for females than for males. The number of deaths in 1950–54, expressed as a percentage of those in 1930–34, was 113 for males and 148 for females.

TABLE XI
DEATHS FROM SUICIDE, BY SEX

Method of Suicide	Males		Females	
	Number	Percentage	Number	Percentage
Analgesics and Soporific Drugs ..	222	7·0	312	16·7
Domestic Gas ..	1,356	42·7	1,072	57·5
Hanging	612	19·3	128	6·9
Drowning	290	9·1	178	9·5
Cutting and Piercing	160	5·0	27	1·4
Firearms	197	6·2	6	0·3
All Other Forms ..	341	10·7	142	7·6
All Methods ..	3,178	100·0	1,865	100·0

The rise in the number of deaths since the introduction of gas for domestic use is shown in Table XII. The trend of the deaths during the war years was different for the two sexes. In 1940–44 the number of suicides in males from coal gas poisoning as a percentage of the number in the preceding quinquennium was 68; the same is true for all methods of suicide. The number of suicides in females in 1940–44 was 82 per cent. of the number in 1935–39, but the number of suicides from coal gas poisoning fell only to 87 per cent. The total number of suicides in 1915–19 as a percentage of those in 1910–14 was 75 for males and 97 for females, but the corresponding percentages for coal gas poisoning were 96 for males and 200 for females.

TABLE XII
NUMBER OF SUICIDES BY DOMESTIC GAS POISONING IN ENGLAND AND WALES

Quinquennium	Males	Females
1890–94	2	1
1895–99	8	5
1900–04	43	10
1905–09	268	94
1910–14	432	169
1915–19	419	340
1920–24	1,529	864
1925–29	3,185	1,868
1930–34	5,174	2,983
1935–39	5,482	3,644
1940–44	3,715	3,182
1945–49	4,681	4,081
1950–54	5,833	4,427

Deaths from analgesics and soporific drugs were mainly due to two groups of substances: barbituric acid and its derivatives, and aspirin and salicylates. The number of deaths from both types of drug has rapidly increased. In the 10 years following 1940 they produced approximately the same number of deaths, but since 1950 sleeping draughts and tablets have become almost three times as frequent as aspirin. The number of deaths registered in England and Wales from these substances during recent years is shown in Table XIII. The rate of increase during the period was larger in males than in females. For barbituric acid the number of deaths in 1950–54 was 8·4 times the number in 1940–44 for males and 6·9 times for females. The corresponding ratio for aspirin and salicylates was 4·0 in males and 2·3 in females.

TABLE XIII
DEATHS FROM DRUGS, BY SEX

Quinquennium	Barbituric Acid and Derivatives		Aspirin and Salicylates	
	Males	Females	Males	Females
1940–44	79	132	64	154
1945–49	218	345	189	328
1950–54	664	907	256	357
Total	961	1,384	509	839

HOMICIDE

The death rate from homicide (Table XIV) has decreased steadily during the past 100 years although the improvement has been slow during the past 15 years. In the earlier years the death rate in males

TABLE XIV
HOMICIDE, DEATH RATES PER MILLION, BY SEX

Period	Males	Females
1863–70	26	15
1871–80	19	12
1881–90	15	10
1891–00	12	9
1901–10	10	9
1911–20	8	8
1921–30	7	6
1931–40	6	6
1941	5	5
1942	7	6
1943	6	5
1944	5	5
1945	8	6
1946	5	5
1947	5	6
1948	6	5
1949	4	5
1950	5	5
1951	5	4
1952	5	4
1953	4	4
1954	5	5

was 50 per cent. greater than that in females. A greater improvement has occurred in males than in females; the chance of being a victim of murder or manslaughter is now about equal for the two sexes.

The worst recent record was that of 1945; the deaths rose from 68 in 1944 to 121 in 1945 in males, and from 103 to 131 in females. The deaths in males showed an increase in each age group; in females, however, the deaths at ages under 1 year and at ages 15–24 were each 10 less than in 1944 (Figure).

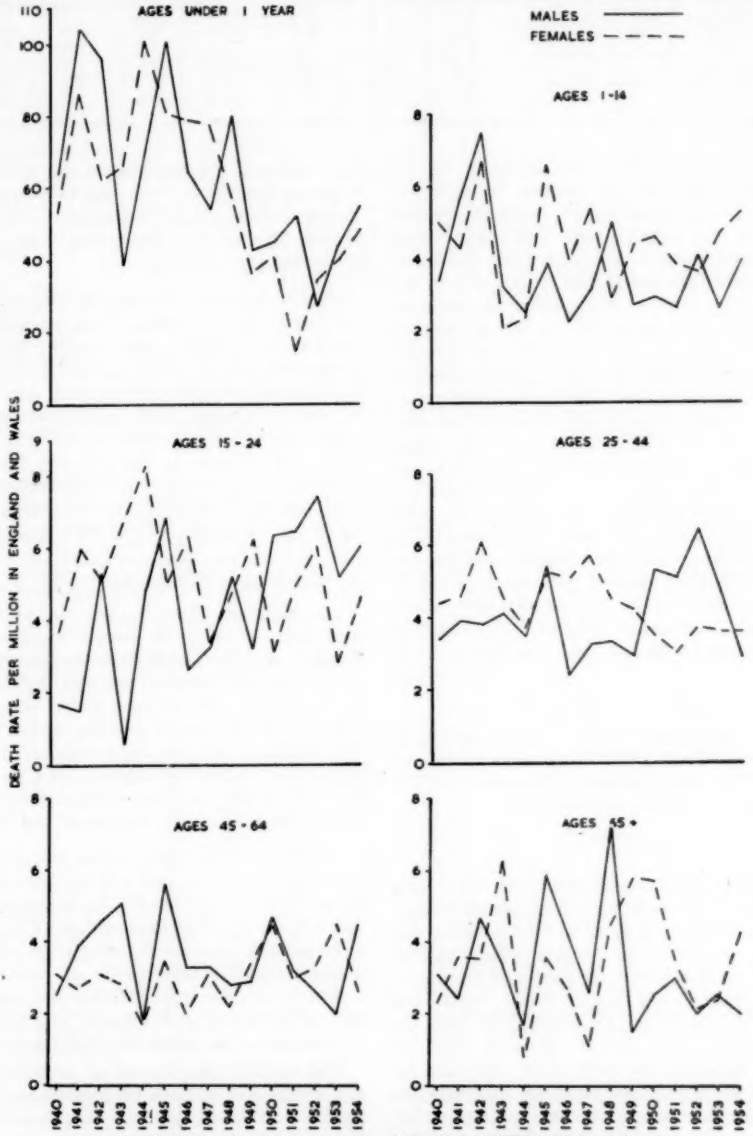

FIGURE.—Homicide. Death rates per million in England and Wales, by age and sex.

TABLE XV

PERCENTAGE DISTRIBUTION OF DEATHS FROM HOMICIDE, BY AGE AND SEX

Sex	Quinquennium	Age (yrs)									No. of Deaths
		Under 1	1–	5–	15–	25–	45–	65–	75–	Total	
Males	1940–44	27·3	10·7	10·7	6·2	20·4	18·5	4·8	1·4	100	421
	1945–49	28·1	7·3	8·2	9·4	20·6	17·6	6·0	2·8	100	466
	1950–54	15·8	6·4	9·0	15·2	31·4	17·4	3·9	1·0	100	488
Females	1940–44	20·3	6·9	7·8	16·8	28·5	12·8	4·6	2·4	100	548
	1945–49	20·9	8·7	7·5	13·0	28·3	13·4	5·5	2·6	100	583
	1950–54	12·2	10·5	9·7	12·6	23·3	21·0	6·6	4·1	100	485

The percentage distribution of deaths by age for the last three quinquennial periods is shown in Table XV. The distributions of deaths in 1940–44 and 1945–49 were very similar. The most striking feature of these years is that almost one-quarter of the deaths were those of infants; in all, 479 babies were killed. The risk of being killed in adult life was differently distributed in the sexes, females being killed at younger ages than males. In 1950–54, the homicide rate underwent considerable changes. A large fall occurred in the death rate under age 1: the actual number of male infants killed fell from 131 in 1945–49 to 77 in 1950–54, and the number of female infants killed fell from 122 to 59. The age distribution of the sexes also changed; in the last quinquennium the females were killed at an older average age than the males. Despite the large fall in the total number of deaths in females, the number of females killed at ages 45–54 rose from 78 in 1945–49 to 102 in 1950–54. The number of deaths in males at ages 15–24 rose from 44 in 1945–49 to 94 in 1950–54; at ages 25–44 they rose from 96 to 153.

REGIONAL DISTRIBUTION

The death rates from motor accidents and from all other accidents in 1950–54 by age and sex are shown in Table XVI (opposite). They are classified by region and by density of population.

(1) MOTOR ACCIDENTS

(*a*) *Regional Classification.*—In males the largest regional differences were in the age group under 5 years. The lowest death rates in both sexes in this group were in the three regions which include London and the South East, the East, and the South; the highest were in the North, North West, and Wales I. The rates in the former regions were less than half those in the latter. At older ages the position is different. In both sexes the rate for London and the South East remained relatively low until old age. In males, however, mortality after childhood was relatively high in the South and East; this tendency appears also in the females but the rates are less regular. At ages 15–24 and 25–44 the highest death rate in males was in Wales II; this region, however, had the lowest death rate but one in the age group 45–64.

(*b*) *Classification by Population Density.*—In males, the death rates at ages 15–64 increased with decreasing population density and the two lowest age groups showed the same tendency. The difference between the mortality in the rural and urban areas was largest at ages 15–24 and 25–44, the rural rate being almost twice that of the lowest urban rate in these age groups. In females also the death rate showed the same tendency to increase with decreasing population density, but with more irregularity up to age 64; after that age the trend was reversed.

(2) ALL OTHER ACCIDENTS

(*a*) *Regional Classification.*—In both sexes, at ages under 5 years, the lowest death rates were in London and the South East, and the South West, and the highest were in the two regions of Wales. In both sexes also, at ages over 5 years, low rates were found in London and the South East, and in the Midland, and Southern regions. In males, at ages over 5 years, the highest rates were in Wales and the North; in females, they were in the East and West Ridings of Yorkshire and in Wales II.

(*b*) *Classification by Population Density.*—In males, the death rates increased with decreasing population density up to age 65. In this respect, motor accidents and all other accidents are similar. The resemblance is particularly noteworthy at ages 15–24 and 25–44, where there is a large excess of rural over urban mortality in both cases. In females no regular pattern can be discerned after infancy, and the death rate in rural areas was relatively high up to middle age.

The regional distribution of suicides by age and sex is shown in Table XVII (opposite). The relative positions of the regions vary in each age group. The mortality was consistently high in each age group in

TABLE XVI

REGIONAL DISTRIBUTION OF ACCIDENTS, BY AGE AND SEX, IN ENGLAND AND WALES, 1950–54, DEATH RATES PER MILLION

Type of Accident	Region	Males								Females							
		Under 5	5–	15–	25–	45–	65–	75 and Over	All Ages	Under 5	5–	15–	25–	45–	65–	75 and Over	All Ages
Motor Vehicle	Northern	176	102	186	138	134	221	545	158	72	49	31	17	39	87	131	43
	East and West Riding	98	86	234	131	136	254	458	152	50	37	29	18	45	95	193	44
	North Western	134	97	211	127	133	227	628	154	74	44	24	22	42	107	236	50
	North Midland	111	106	267	151	147	242	359	167	65	41	35	17	36	82	190	44
	Midland	92	90	267	163	139	332	593	174	62	49	35	27	49	124	201	53
	Eastern	58	88	293	167	152	254	407	174	48	37	49	24	47	83	157	48
	London and S.E.	46	80	206	108	110	214	562	131	38	31	36	18	40	102	245	47
	Southern	64	70	292	183	115	292	398	183	42	38	57	21	32	89	142	45
	South Western	76	92	309	147	144	209	254	164	54	33	44	12	44	84	134	43
	Wales I	138	91	190	140	131	180	462	155	70	54	45	19	34	93	162	46
	Wales II	68	73	357	206	124	243	388	185	57	34	48	16	29	95	80	39
	All Conurbations	86	82	185	112	115	229	617	134	51	35	33	19	44	112	259	49
	Urban 1	96	87	185	117	113	263	455	136	62	40	43	16	40	107	190	47
	Urban 2	97	100	205	125	118	243	481	146	65	38	24	16	39	89	166	43
	Urban 3	95	83	246	137	129	209	432	151	56	42	33	17	30	84	159	41
	Rural Areas	98	105	357	216	191	276	385	217	58	45	48	27	50	81	136	50
All Others	Northern	415	116	285	328	435	589	2,077	385	310	41	22	31	92	415	3,021	205
	East and West Riding	343	98	192	195	288	594	2,492	303	279	37	28	29	93	504	3,225	228
	North Western	391	131	168	170	268	516	2,058	279	291	42	27	30	90	437	2,687	204
	North Midland	340	96	245	234	294	467	1,589	293	249	32	21	25	97	296	2,439	190
	Midland	380	108	153	134	225	511	1,927	241	264	34	18	25	83	429	2,700	188
	Eastern	344	95	265	192	207	419	1,615	269	266	33	51	41	87	386	2,366	221
	London and S.E.	248	80	163	133	203	427	1,682	220	190	27	26	32	87	350	2,130	184
	Southern	352	81	184	158	203	384	1,507	236	240	24	24	27	77	328	2,346	205
	South Western	335	107	206	207	256	443	1,443	275	187	41	33	26	88	302	1,973	193
	Wales I	464	120	262	357	470	680	1,769	407	322	49	27	25	94	420	2,671	193
	Wales II	418	131	283	265	354	456	1,612	345	414	30	26	35	110	341	2,280	227
	All Conurbations	305	97	150	145	229	498	1,960	240	224	33	23	31	89	410	2,436	186
	Urban 1	315	104	168	153	243	521	1,937	256	229	39	23	26	81	397	2,633	196
	Urban 2	343	101	187	192	262	456	2,045	281	285	31	41	28	96	403	2,574	223
	Urban 3	365	106	187	211	302	516	1,729	298	286	35	25	31	97	411	2,507	218
	Rural Areas	419	105	298	274	312	428	1,511	325	278	35	35	31	81	324	2,259	195

Urban 1 = Areas with populations of 100,000 and over. Urban 2 = Areas with population of 50,000 and under 100,000.
Urban 3 = Areas with populations of under 50,000.

TABLE XVII

REGIONAL DISTRIBUTION OF SUICIDES, BY AGE AND SEX, ENGLAND AND WALES 1950–54. DEATH RATES PER MILLION

Region	Males						Females					
	15–	25–	45–	65–	75 and Over	15 and Over	15–	25–	45–	65–	75 and Over	15 and Over
Northern	41	109	272	399	416	182	13	49	111	111	67	69
East and West Riding	48	104	274	386	606	192	14	49	147	185	88	92
North Western	44	108	279	449	543	197	18	53	148	200	155	100
North Midland	31	100	280	441	423	185	8	58	144	183	92	91
Midland	43	97	260	500	513	181	18	62	160	197	154	102
Eastern	44	99	273	383	389	180	11	56	150	160	107	93
London and S.E.	51	108	251	400	474	183	18	63	145	173	121	97
Southern	50	104	232	395	353	164	14	58	141	170	128	94
South Western	40	117	253	402	288	177	14	55	137	170	98	91
Wales I	31	73	214	326	292	142	18	50	122	81	120	74
Wales II	41	108	201	331	328	158	22	35	117	101	40	68
All Conurbations	47	109	259	430	542	187	17	62	146	186	130	97
Urban 1	54	106	258	431	479	186	18	63	179	204	152	112
Urban 2	39	104	278	485	523	197	8	51	164	201	133	102
Urban 3	41	99	267	406	429	185	17	55	141	165	109	92
Rural Areas	39	98	248	350	292	160	13	44	103	111	68	66

Urban 1 = Areas with population of 100,000 and over. Urban 2 = Areas with population of 50,000 and under 100,000.
Urban 3 = Areas with population of under 50,000.

the North West region, and was low in both regions of Wales. At ages 75 and over, in both sexes, the lowest regional rate was less than half the highest regional rate. In both sexes and in each age group (with the exception of females aged 15–24), the suicide rate was lower in rural than in urban areas.

Seasonal Distribution

The seasonal distribution of all violent deaths for the past 15 years is shown in Table XVIII. There is a winter maximum and a summer minimum. The difference between winter and summer was more pronounced in the quinquennium including the war years than in the two following quinquennia. This seasonal analysis gives only a composite picture. An examination of separate causes is a necessary preliminary to any search for explanation.

Suicide.—The suicide rate is highest in summer. Explanations of the fall in the number of suicides during the more gloomy weather require more than statistical data alone. It is interesting to recall that low rates were recorded during the two world wars, but it is the merest speculation to compare psychological and meteorological gloom.

Motor Vehicles.—The maximum is in winter. This is, perhaps, a little surprising, since traffic is much increased in summer, especially at week ends by private cars. There must, it seems, be some connexion here with the more dangerous driving conditions which prevail in the winter months.

Accidental Falls.—There is a high peak in winter. Deaths from falls at home or in residential institutions made up 74 per cent. of all deaths from falls during 1950–54; furthermore, 80 per cent. of deaths from falls during 1950–54 occurred in the age group 65 and over. Thus the weather cannot provide a direct explanation. Nevertheless, more time is spent indoors in winter and it may be that among old people there is an increased risk of death, in cold weather, from complications following an accident.

Other Accidents.—These also show a summer minimum and winter maximum. The seasonal variation in the last two quinquennia was not as large as that found for motor vehicles and accidental falls.

Summary

During the 15 years 1940–54, changes occurred in the age and sex distribution of violent deaths. Death rates in childhood declined continuously, the rate in 1954 being only one-half that of 15 years earlier. At ages over 15 years the death rate fell to a minimum in the late 1940s; it rose subsequently but the rate at the end of the period was still below that at the beginning.

In males, the smallest improvement in the death rate from accidents was at ages 15–19 and 20–24. The death rate from motor accidents has declined in each age group in both sexes, except among males aged 15–24, where the death rate in 1954 was slightly above that of 15 years earlier. Deaths of motor cyclists and their passengers figure prominently in these age groups.

A large improvement occurred in the death rates from falls during the period, especially in childhood. In males the death rate was much larger than in females until age 54; at ages 65–70 and 75 and over, the chances of a fatal fall was greater for females. More males than females below the age of 45 had a fatal fall at home; after this age the females were in excess. The death rate from burns was larger among females than males, particularly in childhood and old age.

Table XVIII

SEASONAL DISTRIBUTION OF DEATHS, ENGLAND AND WALES, 1940–54

(Average number of deaths per day expressed as a percentage of the annual average)

Type of Violence	Quin-quennium	Month												
		Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.	Total
All except War Casualties	1940–44	121	112	106	98	92	93	89	88	90	94	101	117	100
	1945–49	108	106	104	100	97	97	102	95	93	93	100	106	100
	1950–54	108	108	101	102	95	95	97	95	93	98	99	110	100
Suicide	1940–44	91	104	112	110	106	117	106	95	93	89	84	93	100
	1945–49	93	93	107	111	107	111	107	93	96	89	98	95	100
	1950–54	97	100	105	113	110	109	99	96	91	95	92	93	100
Motor Vehicle	1940–44	109	96	96	87	80	78	80	88	101	112	127	147	100
	1945–49	95	86	77	89	92	96	103	102	103	113	119	124	100
	1950–54	93	82	83	89	93	94	103	105	104	112	117	125	100
Falls	1940–44	133	129	109	102	90	81	79	83	89	91	99	115	100
	1945–49	122	127	114	97	94	88	88	85	90	87	97	112	100
	1950–54	120	123	106	95	87	88	88	87	93	99	99	117	100
Other Accidents	1940–44	135	115	107	97	94	97	92	87	84	88	94	110	100
	1945–49	115	114	111	99	95	93	105	100	87	87	93	100	100
	1950–54	116	122	106	106	92	92	98	94	87	90	91	107	100

There was considerable variation in the death rate from accidents between different regions and between different areas of population density. The rate decreased with increasing population up to age 64. At ages 15–44, the death rate from motor vehicle accidents in rural areas was twice that in urban areas. In old age the death rate from motor accidents was largest in the cities and the lowest in rural areas.

At ages over 45, the suicide rate declined during the war but has risen since 1945. The rise has been greatest among the older women, where the rate in 1954 was almost double that of earlier years. Analgesics and soporific drugs have become a numerically important method of suicide, the only one in which the number of deaths in females exceeded those in males; the use of these drugs, however, is increasing faster among men than among women. Domestic gas is still somewhat more frequently employed as a means of suicide. The death rate from hanging has remained fairly constant during the period, and the more disfiguring forms of suicide have become less frequent.

In the first 10 years, the ratio of female to male victims of homicide was five to four; in 1950–54 the rate in females declined and the sex ratio became approximately unity. A large decrease in the number of homicides under 1 year of age was recorded in 1950–54; in the same period increases were recorded for females at ages 45–54 and males at ages 15–44.

The trend of violent deaths suggests that the number of fatal accidents among pre-school children could be considerably reduced by better supervision on the part of adults. The very large accident rate at home among the elderly could also be reduced since many of these fatalities were the result of a fall on the same level, such as might be due, for example, to polished floors or uneven floor covering. The rising rate of suicide among women over age 65 could be a reflection of modern social conditions.

I am indebted to Miss B. Hafner and Miss W. Williams for most of the rather laborious computations.

REFERENCE

Swinscow, D. (1953). *Arch. Dis. Childh.*, **28**, 26.

Brit. J. prev. soc. Med. (1957), **11**, 102–105

AN EXAMINATION OF FERTILITY OF WOMEN FOLLOWING PREGNANCY ACCORDING TO HEIGHT AND WEIGHT

BY

THOMAS McKEOWN AND R. G. RECORD

Department of Social Medicine, University of Birmingham

Although there is reason to believe that differential fertility of biological origin helps to maintain genetic stability, it is by no means easy to obtain acceptable evidence of this phenomenon in man. There are at least two difficulties. In the first place it is technically difficult to record the subsequent fertility of individuals on whom a character such as intelligence, stature, weight etc. has been measured. What is usually noted is the correlation between a character and the number of sibs, an observation much more difficult to interpret. Secondly, the variation in fertility to be expected on theoretical grounds is such that very large numbers of observations would probably be required to establish it. Some objections to conclusions based upon metrical characters such as birth weight have been discussed elsewhere (McKeown, 1956).

An investigation of approximately a thousand representative women during and after pregnancy provided an opportunity to observe the incidence of conception in the two years after delivery. The observation has been used as an index of fertility, and this report examines its relation to maternal age, parity, height, and weight, and to an index of body build.

METHODS

The observations were derived from an investigation of pregnancies of all mothers, domiciled in the County Borough of Smethwick, whose children were born between April 1, 1949 and March 31, 1950. A full account of the data has been given previously (McKeown and Record, 1957a), and here we shall refer only to those features related to the examination of fertility.

(*a*) *Height and Weight of Mother.*—These were measured at the first antenatal examination, and the weight also at 3, 6, 9, 12, and 24 months after delivery. (An index of body build was derived from the observed regression of the weight at 3 months on height.) Women were divided into five groups according to whether they were:

A. More than 15 per cent. below expected weight.

B. 5–15 per cent. below expected weight.

C. Within 5 per cent. of expected weight.

D. 5–15 per cent. above expected weight.

E. More than 15 per cent. above expected weight.

(*b*) *Fertility.*—After examination of various indices of fertility, the incidence of conception within 2 years after delivery was considered to be the most satisfactory for the present purpose. For the first 12 months of this period women recorded their menstrual cycles in a diary and the date of the last menstrual period preceding pregnancy is known. Information about pregnancy in the second year was obtained by interrogating the mother at 24 months, and the date of onset is therefore less exactly known. A few conceptions in the last months were undoubtedly missed by women who were not aware that they were pregnant when seen at 24 months.

1,266 mothers gave birth to single liveborn children in the 12-month period (April, 1949–March, 1950), but 39, whose infants died in the first year after birth, were excluded. Antenatal, birth, and postnatal records were available for 1,118 of the remaining 1,227 women, and age and parity were known for 1,110. However, 171 of these women were untraced 2 years after delivery, and the observations which follow are based on the remaining 939.

Results

195 (20·8 per cent.) of the 939 women traced at 2 years after delivery became pregnant within that period, and five of these conceived twice. Fig. 1 shows the mean number of pregnancies known to have begun before each month until the 15th, and a straight line has been fitted to the points. After 15 months the date of onset of gestation is less accurately known, and since, for obvious reasons, no pregnancies were recorded in the last 2 months, the records are evidently reasonably complete only until 22 months. The proportion of women who had conceived by 22 months (20·8 per cent.) is shown in Fig. 1, and the straight line fitted to the earlier data has been extended (by a dotted line).

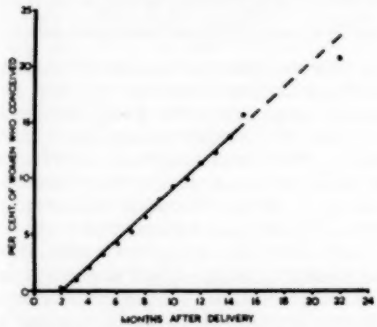

Fig. 1.—Cumulative incidence of conception within 24 months of delivery.

Since it is quite likely that a few early pregnancies were missed in the months immediately before the 22nd, the total cumulative incidence at that time is probably a little low. It is concluded that the cumulative proportions of women who became pregnant in each month after birth fall on a straight line, at least until the 22nd month.

Table I gives the incidence of conceptions within 24 months according to age and parity. The relationship is quite consistent: within each parity the proportion of women who became pregnant declines with age; within each age group after the first (under 25) the trend with parity is U-shaped, incidence being lowest for women who had two pregnancies (including the one just completed at the beginning of the 2-year period). The relationship is examined further in Fig. 2, which gives the cumulative incidence of conceptions for two age groups within each parity. In each case the frequency of conception is greater for the younger women.

Table I

INCIDENCE OF CONCEPTIONS (PER CENT.) WITHIN 24 MONTHS AFTER DELIVERY ACCORDING TO AGE AND PARITY

Parity	Age (yrs)								All Ages	
	Under 25		25–29		30–34		35 and Over			
	No.	Per cent.	No.	Per cent.	No.	Per cent.	No.	Per cent.	No.	Per cent.
1	184	25·0	121	17·4	48	16·7	9	11·1	362	21·0
2	76	31·6	109	13·8	56	10·7	31	6·5	272	17·3
3 and Over	20	50·0	83	26·5	95	22·1	107	17·8	305	23·6
All Parities	280	28·6	313	18·5	199	17·6	147	15·0	939	20·8

The examination of fertility according to height and weight is based on 918 (of the 939) women; weight or height was unknown for the remaining 21, of whom five became pregnant within the 2 years.

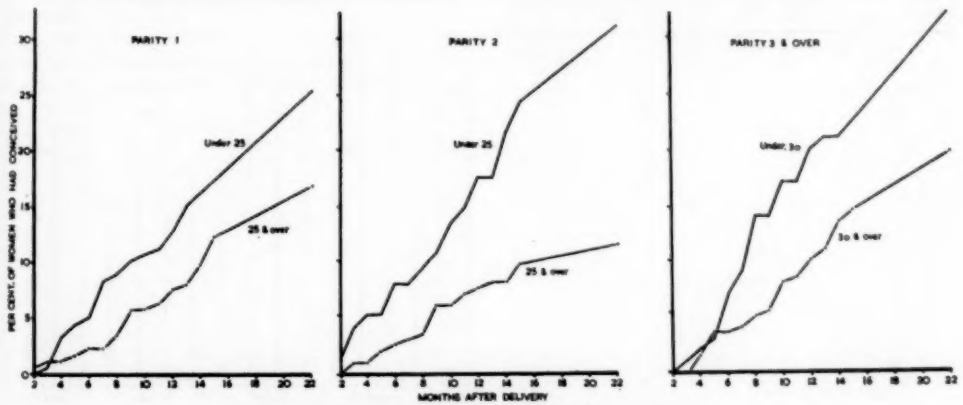

Fig. 2.—Cumulative incidence of conception within 24 months of delivery, related to age and parity.

The incidence of conception is given in Table II for five weight groups, and, because of the marked association between fertility and age and parity, the data have been standardized to remove the influence of these variables. The only noteworthy association is the low incidence of pregnancy in women in the heaviest group (170 lb. and over).

TABLE II

FERTILITY* RELATED TO WEIGHT

Weight 3 mths after Delivery (lb.)	No. of Women	No. who Conceived	Percentage who Conceived	
			Crude	Standardized†
Under 110	121	24	19·8	21·0
110–129	301	70	23·3	22·9
130–149	292	64	21·9	22·3
150–169	131	22	16·8	19·0
170 and Over	73	10	13·7	11·1
Total ..	918	190	20·7	—

* In this and subsequent Tables, fertility refers to the incidence of conception within 24 months after delivery.
† To remove influence of age and parity.

The relationship to height (Table III) is less marked, and on the number of observations available there is little evidence that fertility is related to this variable.

TABLE III

FERTILITY RELATED TO HEIGHT

Height (in.)	No. of Women	No. who Conceived	Percentage who Conceived	
			Crude	Standardized*
Under 62	163	36	22·1	22·9
62 and 63	243	56	23·0	22·8
64 and 65	306	58	19·0	19·4
66 and Over	206	40	19·4	19·0
Total ..	918	190	20·7	—

* To remove influence of age and parity.

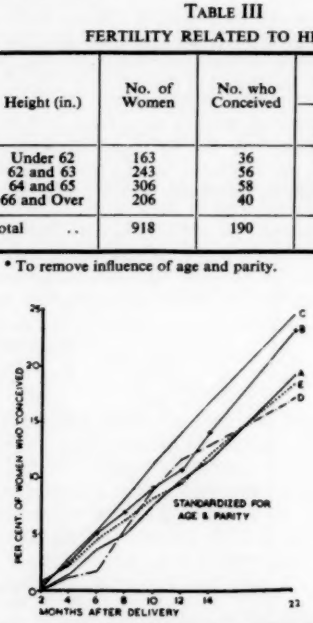

FIG. 3.—Cumulative incidence of conception within 24 months of delivery, related to body build.

Fertility was also examined (Table IV) in relation to body build, by using the five sub-groups referred to above. The incidence of conception is greatest (24·7 per cent. after standardization) in women within 5 per cent. of the expected weight for height.

TABLE IV

FERTILITY RELATED TO BODY BUILD

Body Build Group		A	B	C	D	E	Total
No. of Women		146	230	247	142	153	918
No. of Women who Conceived ..		26	51	60	26	27	190
Incidence of Conceptions (per cent.)	(a) Crude ..	17·8	22·2	24·3	18·3	17·6	20·7
	(b) Standardized for Age and Parity ..	19·2	23·3	24·7	17·4	18·5	—

The data are shown as cumulative percentages at 2-monthly intervals after delivery in Fig. 3.

In a previous examination of the relation between body build and weight changes associated with pregnancy (McKeown and Record, 1957c) it was noted that there was a considerable difference in the experience of women according to their height. It therefore seemed worthwhile to explore the association between fertility and body build, while also taking account of height. For this purpose, because of small numbers it has been necessary to reduce the five groups to three, by combining the two groups more than 5 per cent. below the expected weight for height (A and B) and the two groups more than 5 per cent. above it (D and E). The data are given separately for women in four height groups, and the results have again been standardized to correct for the influence of age and parity (Fig. 4).

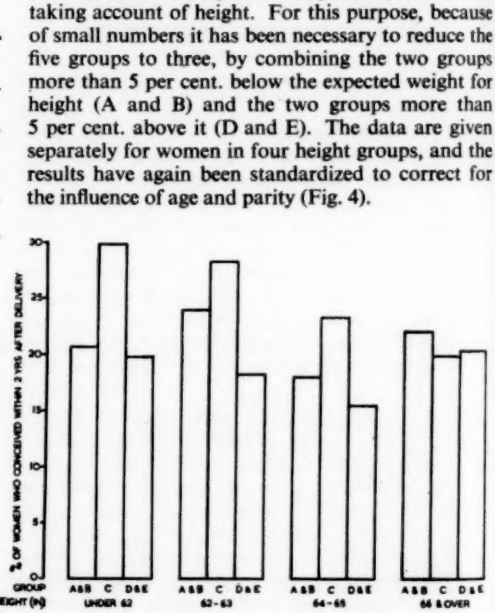

FIG. 4.—Fertility related to height and body build.

In three height groups (under 62, 62–63, and 64–65 in.), the frequency of conception is highest among women within 5 per cent. of the expected weight for height; in the fourth height group (66 in. and over) there is little difference between the three body build groups.

Discussion

It should be said at once that, for assessment of the relationship between fertility and variables such as height, weight, and body build, neither the present population of women, nor the index of fertility employed, is beyond criticism. In the first place, being restricted to women who had completed at least one pregnancy, the data exclude infertile women, and it is possible that the relationship would be different if these were included. Secondly, the frequency of conception within 2 years of delivery is not an ideal index of fertility. We are fairly satisfied that the results are little affected by the duration of lactation, which was shown to be unrelated to weight or body build (McKeown and Record, 1957b). A more serious objection is that we have no information about the influence of contraception, and data examined by Gioiosa (1955) suggest that the frequency of conception within 2 years of delivery is much higher than 20 per cent. (the level here recorded) in a group of women among whom reproduction is unrestricted. Whether the use of contraceptives affects the relationship of fertility to variables such as height and weight is an open question, but the fact that well-to-do women are taller and have smaller families than poor women suggests that it might. On the other hand, the population of women was drawn from a working-class area in which variation in social circumstances was small.

In spite of these reservations it seemed well worth while to take the opportunity provided by a careful record of menstruation following pregnancy, to examine the relationship between fertility and height, weight, and body build. On the numbers of observations available the results must be interpreted cautiously, but the evidence in Fig. 4 suggests that fertility may be somewhat greater in women who are close to the average body build than in those who are fat or thin. It should be noted that if this conclusion is accepted—and the data are regarded as no more than suggestive—it scarcely justifies a generalization about the relationship between fertility and other continously distributed characters such as height. For it seems probable that the incidence of pathological conditions associated with low fertility is higher in fat or thin women than in tall or short women.

Summary

The incidence of conception within 2 years after delivery of a previous child (20·8 per cent.) was used as an index of the fertility of 939 women. Fertility declined with age; the trend in association with parity was U-shaped, incidence of conception being lowest in women who had had two previous pregnancies. The only evidence of association between fertility and weight or height was a low incidence of conception (11·1 per cent.) among the 73 women in the heaviest weight group (170 lb. and over).

The observed regression of weight on height was used to divide women into five groups according to body build. The data suggest that fertility is somewhat higher among women within 5 per cent. of the expected weight for height than in those above or below this level.

References

Gioiosa, R. (1955). *Amer. J. Obstet. Gynec.*, 70, 162.
McKeown, T. (1956). Proceedings of the First International Congress of Human Genetics, Copenhagen, August, 1956. *Acta genet. statist. med.* (in the press).
—— and Record, R. G. (1957a). *J. Endocrinol.*, 15 (in the press).
—— —— (1957b). *Ibid.*
—— —— (1957c). *Ibid.*

CORRIGENDA

The paper entitled "Occurrence and Aetiology of Lung Cancer in Norway in the Light of Pathological Anatomy" by Leiv Kreyberg, which appeared in the October (1956) issue of the *Journal*, vol. 10, p. 145, contains two errors:

(1) In the key to Figure 7 (p. 154) the letters "P" and "C" have been transposed. The same error appears in the same author's paper on "Lung Cancer and Tobacco Smoking in Norway" published in the *British Journal of Cancer* (1955), **9**, 495 (Diagram p. 504), from which the illustration in this *Journal* was taken. The figures which form the basis of the diagram are given in Table III of the Appendix in the *British Journal of Cancer*, **9**, 510.

(2) Table XII (p. 153) should read as follows:

OCCUPATION OF MALES WITH GROUP I TUMOURS BY SMOKING LEVEL

Occupational Category	Smoking Level (g.)										Ratio I:II
	0	4	9	14	19	24	29	49	50+	Total	
"Open Air"	3	2	13	24	5	6	5	2	2	62	1:2·9
"Clerical and Professional"			6	11	15	5	4	10	2	53	1:5·9
"Dusty" ..		2	22	43	9	9	6	6	1	98	1:6·5

Two errors in Table XII were caused by transferring two "waiters" from the "Clerical and Professional" category to the "Dusty" category. Accordingly, in the text (on p. 154, l. 14) one should read 98 instead of 92, and (on l. 18) six instead of eight.